GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:27 ; Search time 25.6438 Seconds
(without alignments)
954.610 Million cell updates/sec

Title: US-09-079-819-178
Perfect score: 4321
Sequence: 1 MILQAHLHSLCLLMLYLATG.....DKGKDNVESAQASEVKPLRS 832

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4321	100.0	832	1	US-08-431-560-1	Sequence 1, Appli
2	4321	100.0	832	1	US-08-463-345-1	Sequence 1, Appli
3	696	16.1	913	2	US-08-474-067-6	Sequence 6, Appli
4	696	16.1	913	2	US-08-474-068A-6	Sequence 6, Appli
5	696	16.1	913	2	US-08-472-481-5	Sequence 5, Appli
6	684.5	15.8	916	1	US-08-188-228-48	Sequence 48, Appl
7	684.5	15.8	916	1	US-08-332-643-42	Sequence 42, Appl
8	684.5	15.8	916	1	US-08-332-638-48	Sequence 48, Appl
9	657.5	15.2	713	1	US-08-188-228-62	Sequence 62, Appl
10	657.5	15.2	713	1	US-08-332-643-56	Sequence 56, Appl
11	657.5	15.2	713	1	US-08-332-638-62	Sequence 62, Appl
12	650.5	15.1	555	2	US-08-453-702A-98	Sequence 98, Appl
13	634	14.7	837	2	US-08-474-067-7	Sequence 7, Appli
14	634	14.7	837	2	US-08-474-068A-7	Sequence 7, Appli
15	634	14.7	837	2	US-08-472-481-6	Sequence 6, Appli
16	633	14.6	556	1	US-07-998-003A-98	Sequence 98, Appl
17	633	14.6	556	1	US-08-453-274B-98	Sequence 98, Appl
18	633	14.6	556	1	US-08-453-695A-98	Sequence 98, Appl
19	633	14.6	556	1	US-08-268-161A-98	Sequence 98, Appl
20	633	14.6	556	4	US-09-099-639-98	Sequence 98, Appl
21	630.5	14.6	712	2	US-08-474-067-2	Sequence 2, Appli
22	630.5	14.6	712	2	US-08-474-067-5	Sequence 5, Appli
23	630.5	14.6	712	2	US-08-474-068A-2	Sequence 2, Appli
24	630.5	14.6	712	2	US-08-474-068A-5	Sequence 5, Appli
25	630.5	14.6	712	2	US-08-472-481-2	Sequence 2, Appli
26	630.5	14.6	717	2	US-08-474-067-4	Sequence 4, Appli
27	630.5	14.6	717	2	US-08-474-068A-4	Sequence 4, Appli
28	630.5	14.6	717	2	US-08-472-481-4	Sequence 4, Appli
29	608	14.1	884	2	US-08-474-067-8	Sequence 8, Appli
30	608	14.1	884	2	US-08-474-068A-8	Sequence 8, Appli
31	608	14.1	884	2	US-08-472-481-7	Sequence 7, Appli
32	593	13.7	878	1	US-08-237-919-2	Sequence 2, Appli
33	593	13.7	878	4	US-08-732-429-2	Sequence 2, Appli
34	593	13.7	878	4	US-09-798-267-2	Sequence 2, Appli
35	593	13.7	878	4	US-09-798-267-3	Sequence 3, Appli
36	593	13.7	878	5	PCT-US95-05518-2	Sequence 2, Appli
37	592.5	13.7	117	4	US-09-187-859-25	Sequence 25, Appl
38	574.5	13.3	796	2	US-08-738-349-2	Sequence 2, Appli
39	560.5	13.0	796	2	US-08-738-349-4	Sequence 4, Appli
40	559	12.9	822	2	US-08-474-067-9	Sequence 9, Appli
41	559	12.9	822	2	US-08-474-068A-9	Sequence 9, Appli
42	559	12.9	822	2	US-08-472-481-8	Sequence 8, Appli
43	555	12.8	106	1	US-08-326-117B-6	Sequence 6, Appli
44	555	12.8	106	3	US-08-982-129-6	Sequence 6, Appli
45	554	12.8	106	4	US-09-178-176B-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-08-431-560-1
; Sequence 1, Application US/08431560
; Patent No. 5620855
;  GENERAL INFORMATION:
;    APPLICANT:  Anne H. Dantzig, et al.
;    TITLE OF INVENTION:  Mammalian Influx Peptide
;    TITLE OF INVENTION:  Transporter
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  IN
;      COUNTRY:  U.S.A.
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.0 Mb storage
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/431,560
;      FILING DATE:  01-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/013,462
;      FILING DATE:  04-FEB-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-431-560-1

  Query Match           100.0%;  Score 4321;  DB 1;  Length 832;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  832;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy    181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy    241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy    301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy    361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy    421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy    721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832


RESULT 2
US-08-463-345-1
; Sequence 1, Application US/08463345
; Patent No. 5710018
;  GENERAL INFORMATION:
;    APPLICANT:  Anne H. Dantzig, et al.
;    TITLE OF INVENTION:  Mammalian Influx Peptide
;    TITLE OF INVENTION:  Transporter
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  IN
;      COUNTRY:  U.S.A.
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.0 Mb storage
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh
;      SOFTWARE:  Microsoft Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/463,345
;      FILING DATE:  05-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/013,462
;      FILING DATE:  04-FEB-1993
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  832 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-463-345-1

  Query Match           100.0%;  Score 4321;  DB 1;  Length 832;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  832;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy     61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy    121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy    181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy    241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy    301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy    361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy    421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy    721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832


RESULT 3
US-08-474-067-6
```

```
; Sequence 6, Application US/08474067
; Patent No. 5811518
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,067
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1682
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-474-067-6

  Query Match             16.1%;  Score 696;  DB 2;  Length 913;
  Best Local Similarity   26.9%;  Pred. No. 6.9e-53;
  Matches  209;  Conservative  117;  Mismatches  328;  Indels  124;  Gaps   22;

Qy     126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183
           |  :   :   | ::   |:| | |     |:                  |  :||  :
Db      41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy     184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENSFSDTT 228
                 : | :   |       | :      |:| :|   |           | :|:  :
Db      82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy     229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269
             |    |:|                  |  | |: : |||  | | :: ::|
Db     137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy     270 QYSLVDKEKLPRFPFSI-----DQ----------EGDIYVTQPLDREEKDAYVFYAVAK 313
               |::|     :|:     ||           | : ||:|||||:  ::   | |
Db     190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy     314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373
           |  |  :  |::| : | |:||| |     |    | |  : |  : |:|| | |: |
Db     246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy     374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428
           |  | |||: | |  |   :| |    | :      | ::   || | |: :| :
Db     306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy     429 -FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
                | | |:||  | |   : || :   :  :   :  :  || |:| | :
Db     366 GLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWN 422

Qy     487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
             | :  ||  |:  : |||::| | | : ||:||||  :  :   |||  ||  |:::
Db     423 ARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482

Qy     547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISYSL 602
             | |  :: | ||||:| |  :    :  | :  |: :   ||:||    :   : ||
Db     483 PQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRYSK 542

Qy     603 RGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILM 658
             |   ||||| | |:| : | ||||:     : |    :|:: |   :|      : |:
Db     543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIYLL 602

Qy     659 DVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQND 715
           |:||| |::  |: |    |  | | ::   | | |     || | | |     |:: :
Db     603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657

Qy     716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFR 774
           | : :|:| ||:|| |    |   | | | | | | |       | |  | | : | |
Db     658 WTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC-- 715

Qy     775 PAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
             | :  :   |:  | ::  || | |:| :| : :   | :|    |:   : |
Db     716 -----TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768


RESULT 4
US-08-474-068A-6
; Sequence 6, Application US/08474068A
; Patent No. 5837525
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,068A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1683
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-474-068A-6

  Query Match             16.1%;  Score 696;  DB 2;  Length 913;
  Best Local Similarity   26.9%;  Pred. No. 6.9e-53;
  Matches  209;  Conservative  117;  Mismatches  328;  Indels  124;  Gaps   22;
```

```
Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183
          | :  :   | ::   |:| | |     |:                   |  :||  :
Db     41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy    184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENSFSDTT 228
              : | :   |       | :      |:| :|   |             | :|:  :
Db     82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy    229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269
            |    |:|                  | | |: : |||  | | :: ::|
Db    137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy    270 QYSLVDKEKLPRFPFSI-----DQ-----------EGDIYVTQPLDREEKDAYVFYAVAK 313
            |::|     :|:     ||             | : ||:|||||:  ::   | |
Db    190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy    314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373
          |  |  :  |::| : | |:||| |     |   | |  : |  : |:|| | |: |
Db    246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy    374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428
          |  | |||: | |  |   :| |     | :      | ::   || | |: :| :
Db    306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy    429 -FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
               | | |:||  | |   : || :    :  :   :  :  || |:| | :
Db    366 GLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWN 422

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
          | :  ||  |:  : |||::| | | : ||:||||  :  :   |||  ||  |:::
Db    423 ARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISYSL 602
          | |  :: | ||||:| |  :    :  | :  |: :   ||:||    :   : ||
Db    483 PQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRYSK 542

Qy    603 RGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILM 658
            |   ||||| | |:| : | ||||:     : |    :|:: |   :|      : |:
Db    543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIYLL 602

Qy    659 DVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQND 715
          |:||| |::  |: |    |  |  | ::   | | |     || | | |     |:: :
Db    603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657

Qy    716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFR 774
          | : :|:| ||:|| |    |   | | | | | | |       | |  | | : | |
Db    658 WTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC-- 715

Qy    775 PAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
            | :  :   |:  | ::  || | |:| :| : :   | :|    |:   : |
Db    716 -----TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768


RESULT 5
US-08-472-481-5
; Sequence 5, Application US/08472481
; Patent No. 5863804
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,481
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1686
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  913 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-472-481-5

  Query Match             16.1%;  Score 696;  DB 2;  Length 913;
  Best Local Similarity   26.9%;  Pred. No. 6.9e-53;
  Matches  209;  Conservative  117;  Mismatches  328;  Indels  124;  Gaps   22;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183
          | :  :   | ::   |:| | |     |:                   |  :||  :
Db     41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy    184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENSFSDTT 228
              : | :   |       | :      |:| :|   |             | :|:  :
Db     82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy    229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269
            |    |:|                  | | |: : |||  | | :: ::|
Db    137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy    270 QYSLVDKEKLPRFPFSI-----DQ-----------EGDIYVTQPLDREEKDAYVFYAVAK 313
            |::|     :|:     ||             | : ||:|||||:  ::   | |
Db    190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy    314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373
          |  |  :  |::| : | |:||| |     |   | |  : |  : |:|| | |: |
Db    246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy    374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428
          |  | |||: | |  |   :| |     | :      | ::   || | |: :| :
Db    306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy    429 -FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
               | | |:||  | |   : || :    :  :   :  :  || |:| | :
Db    366 GLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWN 422

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
          | :  ||  |:  : |||::| | | : ||:||||  :  :   |||  ||  |:::
Db    423 ARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISYSL 602
          | |  :: | ||||:| |  :    :  | :  |: :   ||:||    :   : ||
Db    483 PQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRYSK 542

Qy    603 RGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILM 658
            |   ||||| | |:| : | ||||:     : |    :|:: |   :|      : |:
Db    543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIYLL 602

Qy    659 DVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQND 715
          |:||| |::  |: |    |  |  | ::   | | |     || | | |     |:: :
Db    603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657
```

```
Qy    716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFR 774
          | : :|:| ||:|| |    |   | | | | | | |       | |  | | : | |
Db    658 WTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC-- 715

Qy    775 PAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
              |  :   :   |:  | ::  || | |:| :| : :   | :|    |:   : |
Db    716 -----TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768


RESULT 6
US-08-188-228-48
; Sequence 48, Application US/08188228
; Patent No. 5597725
;  GENERAL INFORMATION:
;    APPLICANT:  Suzuki, Shintaro
;    TITLE OF INVENTION:  CADHERIN MATERIALS AND METHODS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 S. Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/188,228
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/049,460
;      FILING DATE:  19 APR 1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,643
;      FILING DATE:  17 APR 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5597725and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31340
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  916 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-188-228-48

  Query Match             15.8%;  Score 684.5;  DB 1;  Length 916;
  Best Local Similarity   29.3%;  Pred. No. 7.3e-52;
  Matches  176;  Conservative  106;  Mismatches  288;  Indels   31;  Gaps   12;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
          |  | |: : |||  | | :: ::|    ||   :||:  |  :: |   |||:    | :
Db    171 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFSINSMSGRM 229

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
          |||:|:|||| :|   | | |  |  :  |:::: | |:||| |   : |    | |
Db    230 YVTRPMDREEHASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNHPEFINQVYNCSVDEG 289

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
           : |  : |:||:| |:  |||  : |||| |||: |   :| | :  | :        :
Db    290 SKPGTYVMTITANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
          :   || : :: :| :               | | |:||    |  | :       : ::
Db    350 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPSEFTASTFAG--EVPENSVE 407

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
          : :  :   | |:| : :   :| || ||  |   | ||| || | | : | :|:|
Db    408 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 467

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
            :     |  ||  |:: :  | |  |: : |:|||| |  :    :: | |  || :
Db    468 FMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPYFPSNHKLIRLEEGVPPGTVLT 527

Qy    587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
            :| ||:      : ||   |   || |:   |:| :|| ||||:     : |    :|
Db    528 TFSAVDPDRFMQQAVRYSKLSDPASWLHINATNGQITTVAVLDRESLYTKNNVYEATFLA 587

Qy    640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFR 699
            : |    |      : |:|:||| | |      :     |:|   :   | | | |
Db    588 ADNGIPPASGTGTLQIYLIDINDNAPELLPKEAQICERPNLNA---INITAADADVHPNI 644

Qy    700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIV 758
          ||: |       ::: :| ::::|| :|:|| |     |   | | | : | | |||
Db    645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPIIVTDSGNPPLSNTS 704

Qy    759 SLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--FIRIKKDKGK 816
            : |  | | :       |    :   |:  | ::  |: | |:| :|  |:   | : |
Db .  705 IIKVKVCPCDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTMVLLFVMWMKRREK 761

Qy    817 D 817
            :
Db    762 E 762


RESULT 7
US-08-332-643-42
; Sequence 42, Application US/08332643
; Patent No. 5639634
;  GENERAL INFORMATION:
;    APPLICANT:  Suzuki, Shintaro
;    TITLE OF INVENTION:  CADHERIN MATERIALS AND METHODS
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,643
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/872,643
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5639634and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30795
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  916 amino acids
;      TYPE:  amino acid
```

```
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-332-643-42

  Query Match             15.8%;  Score 684.5;  DB 1;  Length 916;
  Best Local Similarity   29.3%;  Pred. No. 7.3e-52;
  Matches  176;  Conservative  106;  Mismatches  288;  Indels   31;  Gaps   12;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
          |  | |: : |||  | | :: ::|    ||   :||:  |  :: |   |||:   | :
Db    171 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFSINSMSGRM 229

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
          |||:|:|||| :|   | | |  |  :  |:::::| |:||| |   : |    | |
Db    230 YVTRPMDREEHASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNHPEFINQVYNCSVDEG 289

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
          : |  : |:||:| |:  |||  : |||| |||: |   :| | :  | :      . :
Db    290 SKPGTYVMTITANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
          :  || : :: :| :             | | |:||   |  | :       : ::
Db    350 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPSEFTASTFAG--EVPENSVE 407

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
          : :  :   | |:| : :   :| || ||  |   | ||| || | | : | :|:|
Db    408 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 467

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
            :     |  ||  |:: :  | |  |: : |:|||| |  :    :: | |  || :
Db    468 FMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPYFPSNHKLIRLEEGVPPGTVLT 527

Qy    587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
          :| ||:     : ||   |   || |:   |:| :|| |||:     : |    :|
Db    528 TFSAVDPDRFMQQAVRYSKLSDPASWLHINATNGQITTVAVLDRESLYTKNNVYEATFLA 587

Qy    640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFR 699
          : |    |      : |:|:||| | |      :     |:|   :   | | | |
Db    588 ADNGIPPASGTGTLQIYLIDINDNAPELLPKEAQICERPNLNA---INITAADADVHPNI 644

Qy    700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIV 758
          ||: |       ::: :| ::::|| :|:|| |    |   | | | : | | |||
Db    645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPIIVTDSGNPPLSNTS 704

Qy    759 SLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--FIRIKKDKGK 816
          : |  | | :      |   :   |:  | ::  |: | |:| :|  |:   | : |
Db    705 IIKVKVCPCDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTMVLLFVMWMKRREK 761

Qy    817 D 817
          :
Db    762 E 762


RESULT 8
US-08-332-638-48
; Sequence 48, Application US/08332638
; Patent No. 5646250
;  GENERAL INFORMATION:
;    APPLICANT:  Suzuki, Shintaro
;    TITLE OF INVENTION:  CADHERIN MATERIALS AND METHODS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 S. Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,638
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,643
;      FILING DATE:  17 APR 1992
;      APPLICATION NUMBER:  US/08/049,460
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5646250and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31340
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  916 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-332-638-48

  Query Match             15.8%;  Score 684.5;  DB 1;  Length 916;
  Best Local Similarity   29.3%;  Pred. No. 7.3e-52;
  Matches  176;  Conservative  106;  Mismatches  288;  Indels   31;  Gaps   12;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
          |  | |: : |||  | | :: ::|    ||   :||:  |  :: |   |||:   | :
Db    171 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFSINSMSGRM 229

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
          |||:|:|||| :|   | | |  |  :  |:::::| |:||| |   : |    | |
Db    230 YVTRPMDREEHASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNHPEFINQVYNCSVDEG 289

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
          : |  : |:||:| |:  |||  : |||| |||: |   :| | :  | :      . :
Db    290 SKPGTYVMTITANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
          :  || : :: :| :             | | |:||   |  | :       : ::
Db    350 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPSEFTASTFAG--EVPENSVE 407

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
          : :  :   | |:| : :   :| || ||  |   | ||| || | | : | :|:|
Db    408 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 467

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
            :     |  ||  |:: :  | |  |: : |:|||| |  :    :: | |  || :
Db    468 FMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPYFPSNHKLIRLEEGVPPGTVLT 527

Qy    587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
          :| ||:     : ||   |   || |:   |:| :|| |||:     : |    :|
Db    528 TFSAVDPDRFMQQAVRYSKLSDPASWLHINATNGQITTVAVLDRESLYTKNNVYEATFLA 587

Qy    640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFR 699
          : |    |      : |:|:||| | |      :     |:|   :   | | | |
Db    588 ADNGIPPASGTGTLQIYLIDINDNAPELLPKEAQICERPNLNA---INITAADADVHPNI 644

Qy    700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIV 758
          ||: |       ::: :| ::::|| :|:|| |    |   | | | : | | |||
Db    645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPIIVTDSGNPPLSNTS 704

Qy    759 SLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--FIRIKKDKGK 816
          : |  | | :      |   :   |:  | ::  |: | |:| :|  |:   | : |
Db    705 IIKVKVCPCDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTMVLLFVMWMKRREK 761

Qy    817 D 817
          :
Db    762 E 762
```

```
RESULT 9
US-08-188-228-62
; Sequence 62, Application US/08188228
; Patent No. 5597725
;  GENERAL INFORMATION:
;    APPLICANT:  Suzuki, Shintaro
;    TITLE OF INVENTION:  CADHERIN MATERIALS AND METHODS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 S. Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/188,228
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/049,460
;      FILING DATE:  19 APR 1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,643
;      FILING DATE:  17 APR 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5597725and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31340
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  62:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  713 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-188-228-62

  Query Match             15.2%;  Score 657.5;  DB 1;  Length 713;
  Best Local Similarity   28.7%;  Pred. No. 1.2e-49;
  Matches  199;  Conservative  104;  Mismatches  285;  Indels  105;  Gaps   23;

Qy    160 NGQLYYQIVIQLPMINNVMYFQINNKTGAISL---TREGSQELNPAKNPSYN-----LVI 211
          | :| |::        :  ||::|:  | ::|   |  |      |: |        :::
Db     61 NDKLRYEV--------SSPYFKVNSDGGLVALRNITAVGKTLFVHARTPHAEDMAELVIV 112

Qy    212 SVKDMGGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVEMVENSTDPHPIKITQVRW 264
            ||: |  ::   |: |: |      |:|:        |: : ||   | |  : :|
Db    113 GGKDIQGSLQDIFKFARTSPVPRQKRSIVVS--------PILIPENQRQPFPRDVGKV-- 162

Qy    265 NDPGAQYSLVDKEKLPRFPF-----SIDQE----------GDIYVTQPLDREEKDAYVF 308
                 || ::  |  |      :|||          | : ||: ||||    |
Db    163 ---------VDSDRPERSKFRLTGKGVDQEPKGIFRINENTGSVSVTRTLDREVIAVYQL 213

Qy    309 YAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRD 368
          :    |  || |  |: : | | | ||| |       :  | |    | ::  :|| | |
Db    214 FVETTDVNGKTLEGPVPLEVIVIDQNDNRPIFREGPYIGHVMEGSPTGTTVMRMTAFDAD 273

Qy    369 EENTANSFLNYRIVEQTPKLPMDGLFLIQTYAG-MLQLAKQSLKKQDT---PQYNLTIEV 424
          :  | |: | | | :|||  |   :| |    | :: :   :|  ::|   |:| | ||
Db    274 DPATDNALLRYNIRQQTPDKPSPNMFYIDPEKGDIVTVVSPALLDRETLENPKYELIIEA 333

Qy    425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADE 479
          |               | : | ||  | | | ::      |:  :|  |: :   | |:
Db    334 QDMAGLDVGLTGTATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDD 390

Qy    480 PFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPL 539
          | ||: :  | || |:      : |:| || | : : |||||:| :|   :: | || :||
Db    391 PTTGAWRAAYTIINGNPGQSFEIHTNPQTNEGMLSVVKPLDYEISAFHTLLIKVENEDPL 450

Qy    540 VFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLD-- 597
          |  | |  || |   : | |||| | |          ||:::|: :  | | ||: |
Db    451 VPDVSYGPSSTATVHITVLDVNEGPVFYPDPMMVTRQEDLSVGSVLLTVNATDPDSLQHQ 510

Qy    598 -ISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSE 652
           | ||:  |  ||| |: : | : : | ||||:     | |    :| : |    :
Db    511 TIRYSVYKDPAGWLNINPINGTVDTTAVLDRESPFVDNSVYTALFLAIDSGNPPATGTGT 570

Qy    653 FHLILMDVNDNPPRLAKDYTGLFFCHPLSAP--------GSLIFEATDDDQHLFRGPHFT 704
            : | |||| |          | :|  |          :|  |:| | |     | |
Db    571 LLITLEDVNDNAP----------FIYPTVAEVCDDAKNLSVVILGASDKDLHPNTDP-FK 619

Qy    705 FSLGSGSLQND-WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVT 763
          | :   :: :  |::|||| ||| :|    :  :  | : | : | |:||:  |  | |
Db    620 FEIHKQAVPDKVWKISKINNTHALVSLLQ-NLNKANYNLPIMVTDSGKPPMTNITDLRVQ 678

Qy    764 FCSCVEG--SCFRPAGHQTGIPTVGMAVGILLT 794
          |||      |      :  :|:|     |||:
Db    679 VCSCRNSKVDCNAAGALRFSLPSV-----ILLS 706


RESULT 10
US-08-332-643-56
; Sequence 56, Application US/08332643
; Patent No. 5639634
;  GENERAL INFORMATION:
;    APPLICANT:  Suzuki, Shintaro
;    TITLE OF INVENTION:  CADHERIN MATERIALS AND METHODS
;    NUMBER OF SEQUENCES:  56
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,643
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/872,643
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5639634and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/30795
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  713 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-332-643-56
```

```
  Query Match             15.2%;  Score 657.5;  DB 1;  Length 713;
  Best Local Similarity   28.7%;  Pred. No. 1.2e-49;
  Matches  199;  Conservative  104;  Mismatches  285;  Indels  105;  Gaps   23;

Qy    160 NGQLYYQIVIQLPMINNVMYFQINNKTGAISL---TREGSQELNPAKNPSYN-----LVI 211
          | :| |::        :  ||::|:  | ::|   |  |       |: |         :::
Db     61 NDKLRYEV--------SSPYFKVNSDGGLVALRNITAVGKTLFVHARTPHAEDMAELVIV 112

Qy    212 SVKDMGGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVEMVENSTDPHPIKITQVRW 264
            ||: |  ::   |: |: |      |:|:        |: : ||   | |  : :|
Db    113 GGKDIQGSLQDIFKFARTSPVPRQKRSIVVS--------PILIPENQRQPFPRDVGKV-- 162

Qy    265 NDPGAQYSLVDKEKLPRFPF-----SIDQE----------GDIYVTQPLDREEKDAYVF 308
                  || ::  |  |      :|||           | : ||: ||||    |
Db    163 ---------VDSDRPERSKFRLTGKGVDQEPKGIFRINENTGSVSVTRTLDREVIAVYQL 213

Qy    309 YAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRD 368
          :   |  || |  |: : | | | |||  |       :  | |    | ::  :|| | |
Db    214 FVETTDVNGKTLEGPVPLEVIVIDQNDNRPIFREGPYIGHVMEGSPTGTTVMRMTAFDAD 273

Qy    369 EENTANSFLNYRIVEQTPKLPMDGLFLIQTYAG-MLQLAKQSLKKQDT---PQYNLTIEV 424
          :  | |: | | | :|||  |   :| |    | :: :   :|  ::|   |:| | ||
Db    274 DPATDNALLRYNIRQQTPDKPSPNMFYIDPEKGDIVTVVSPALLDRETLENPKYELIIEA 333

Qy    425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADE 479
          |               | : | ||  | | | ::      |:  :|  |: :   | |:
Db    334 QDMAGLDVGLTGTATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDD 390

Qy    480 PFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPL 539
          | ||: :  | || |:      : |:| || | : : ||||:| :|   :: | || :||
Db    391 PTTGAWRAAYTIINGNPGQSFEIHTNPQTNEGMLSVVKPLDYEISAFHTLLIKVENEDPL 450

Qy    540 VFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLD-- 597
          |  | |  || |   : | |||| | |          ||:::|: :  | | ||: |
Db    451 VPDVSYGPSSTATVHITVLDVNEGPVFYPDPMMVTRQEDLSVGSVLLTVNATDPDSLQHQ 510

Qy    598 -ISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSE 652
          | ||:  |  ||| |: : | : : | ||||:     | |    :| : |    :
Db    511 TIRYSVYKDPAGWLNINPINGTVDTTAVLDRESPFVDNSVYTALFLAIDSGNPPATGTGT 570

Qy    653 FHLILMDVNDNPPRLAKDYTGLFFCHPLSAP--------GSLIFEATDDDQHLFRGPHFT 704
          : | |||||           | :|  |           :|  |:| | |    | |
Db    571 LLITLEDVNDNAP----------FIYPTVAEVCDDAKNLSVVILGASDKDLHPNTDP-FK 619

Qy    705 FSLGSGSLQND-WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVT 763
          | :   :: :  |::|||| ||| :|    :  :  | : | : | |:||:  |  | |
Db    620 FEIHKQAVPDKVWKISKINNTHALVSLLQ-NLNKANYNLPIMVTDSGKPPMTNITDLRVQ 678

Qy    764 FCSCVEG--SCFRPAGHQTGIPTVGMAVGILLT 794
          |||      |      :  :|:|     |||:
Db    679 VCSCRNSKVDCNAAGALRFSLPSV-----ILLS 706


RESULT 11
US-08-332-638-62
; Sequence 62, Application US/08332638
; Patent No. 5646250
;  GENERAL INFORMATION:
;    APPLICANT:  Suzuki, Shintaro
;    TITLE OF INVENTION:  CADHERIN MATERIALS AND METHODS
;    NUMBER OF SEQUENCES:  62
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 S. Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,638
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/872,643
;      FILING DATE:  17 APR 1992
;      APPLICATION NUMBER:  US/08/049,460
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5646250and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31340
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  62:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  713 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-332-638-62

  Query Match             15.2%;  Score 657.5;  DB 1;  Length 713;
  Best Local Similarity   28.7%;  Pred. No. 1.2e-49;
  Matches  199;  Conservative  104;  Mismatches  285;  Indels  105;  Gaps   23;

Qy    160 NGQLYYQIVIQLPMINNVMYFQINNKTGAISL---TREGSQELNPAKNPSYN-----LVI 211
          | :| |::        :  ||::|:  | ::|   |  |       |: |         :::
Db     61 NDKLRYEV--------SSPYFKVNSDGGLVALRNITAVGKTLFVHARTPHAEDMAELVIV 112

Qy    212 SVKDMGGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVEMVENSTDPHPIKITQVRW 264
            ||: |  ::   |: |: |      |:|:        |: : ||   | |  : :|
Db    113 GGKDIQGSLQDIFKFARTSPVPRQKRSIVVS--------PILIPENQRQPFPRDVGKV-- 162

Qy    265 NDPGAQYSLVDKEKLPRFPF-----SIDQE----------GDIYVTQPLDREEKDAYVF 308
                  || ::  |  |      :|||           | : ||: ||||    |
Db    163 ---------VDSDRPERSKFRLTGKGVDQEPKGIFRINENTGSVSVTRTLDREVIAVYQL 213

Qy    309 YAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRD 368
          :   |  || |  |: : | | | |||  |       :  | |    | ::  :|| | |
Db    214 FVETTDVNGKTLEGPVPLEVIVIDQNDNRPIFREGPYIGHVMEGSPTGTTVMRMTAFDAD 273

Qy    369 EENTANSFLNYRIVEQTPKLPMDGLFLIQTYAG-MLQLAKQSLKKQDT---PQYNLTIEV 424
          :  | |: | | | :|||  |   :| |    | :: :   :|  ::|   |:| | ||
Db    274 DPATDNALLRYNIRQQTPDKPSPNMFYIDPEKGDIVTVVSPALLDRETLENPKYELIIEA 333

Qy    425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADE 479
          |               | : | ||  | | | ::      |:  :|  |: :   | |:
Db    334 QDMAGLDVGLTGTATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDD 390

Qy    480 PFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPL 539
          | ||: :  | || |:      : |:| || | : : ||||:| :|   :: | || :||
Db    391 PTTGAWRAAYTIINGNPGQSFEIHTNPQTNEGMLSVVKPLDYEISAFHTLLIKVENEDPL 450

Qy    540 VFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLD-- 597
          |  | |  || |   : | |||| | |          ||:::|: :  | | ||: |
Db    451 VPDVSYGPSSTATVHITVLDVNEGPVFYPDPMMVTRQEDLSVGSVLLTVNATDPDSLQHQ 510

Qy    598 -ISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSE 652
          | ||:  |  ||| |: : | : : | ||||:     | |    :| : |    :
Db    511 TIRYSVYKDPAGWLNINPINGTVDTTAVLDRESPFVDNSVYTALFLAIDSGNPPATGTGT 570

Qy    653 FHLILMDVNDNPPRLAKDYTGLFFCHPLSAP--------GSLIFEATDDDQHLFRGPHFT 704
          : | |||||           | :|  |           :|  |:| | |    | |
Db    571 LLITLEDVNDNAP----------FIYPTVAEVCDDAKNLSVVILGASDKDLHPNTDP-FK 619

Qy    705 FSLGSGSLQND-WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVT 763
          | :   :: :  |::|||| ||| :|    :  :  | : | : | |:||:  |  | |
```

```
Db    620 FEIHKQAVPDKVWKISKINNTHALVSLLQ-NLNKANYNLPIMVTDSGKPPMTNITDLRVQ 678

Qy    764 FCSCVEG--SCFRPAGHQTGIPTVGMAVGILLT 794
          |||      |      :  :|:|     |||:
Db    679 VCSCRNSKVDCNAAGALRFSLPSV-----ILLS 706


RESULT 12
US-08-453-702A-98
; Sequence 98, Application US/08453702A
; Patent No. 5891706
;  GENERAL INFORMATION:
;    APPLICANT:  Suzuki, Shintaro
;    TITLE OF INVENTION:  Protocadherin Materials and Methods
;    NUMBER OF SEQUENCES:  115
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray, &
;      ADDRESSEE:  Borun
;      STREET:  233 South Wacker, 6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/453,702A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5891706and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  32657
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312/474-6300
;      TELEFAX:  312/474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  98:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  555 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-453-702A-98

  Query Match             15.1%;  Score 650.5;  DB 2;  Length 555;
  Best Local Similarity   30.5%;  Pred. No. 3.3e-49;
  Matches  174;  Conservative   87;  Mismatches  252;  Indels   57;  Gaps   13;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQYSLVDKEKLPRFPFSI 286
          |  | |: : |||  | | :: ::| :             |||       :: |   | |
Db      2 WVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA-------DQPPTGIFII 53

Qy    287 DQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVT 345
          :   | : ||:|||||    :   | | |  |  :  |::| : | |:||| |     |
Db     54 NPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVINVIDMNDNRPEFLHQVW 113

Qy    346 VFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQL 405
             | |  : |  : |:|| | |: |  |  | |||: | |  |   :| |    | :
Db    114 NGSVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTPSPNMFTINNETGDIIT 173

Qy    406 AKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIF-EKSDYGNLT 458
             | ::   || | |: :| :             | | |:||  | |   : || :
Db    174 VAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVNDNPPEFTAMTFYGEV- 232

Qy    459 LAEDTNIGSTILTIQATDADEPPTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKP 518
          :  :   :  :  || |:| | :    | |  ||  ||  : |||::| | | : ||
Db    233 --PENRVDVIVANLTVTDKDQPHTPAWNAAYRISGGDPTGRFAILTDPNSNDGLVTVVKP 290

Qy    519 LDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSED 578
          :|||| :  :   |||  ||  |::::   | |  :: | |||| | |: :    :  |
Db    291 IDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNENPYFAPNPKIIRQEEG 350

Qy    579 VAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GS 631
          :  || :  :||:||:     :| |:   |  |||||  | |:| ::| ||||:     :
Db    351 LHAGTMLTTLTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQITTIAVLDRESPYVQNN 410

Qy    632 PYRVQVVATEVGGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSA-------PG 684
          |    :|:: |   :|      : |:|:||| |::           |  |        |
Db    411 IYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL----------PQEAETCETPEPN 460

Qy    685 SLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVL 743
          |:   | | |      ||  |   |   ::: :| ::::|| |:|: :    |   | |
Db    461 SINIAALDYDIDPNAGPFAFDLPLSPVTIKRNWTINRLNGDFAQLNLKIKFLEAGIYEVP 520

Qy    744 IRINDGGRPPLEGIVSLPVTFCSC-VEGSC 772
          | | | | ||   |  | |  | |   | |
Db    521 IIITDSGNPPKSNISILRVKVCQCDSNGDC 550


RESULT 13
US-08-474-067-7
; Sequence 7, Application US/08474067
; Patent No. 5811518
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,067
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1682
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  837 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-474-067-7

  Query Match             14.7%;  Score 634;  DB 2;  Length 837;
  Best Local Similarity   30.4%;  Pred. No. 2e-47;
  Matches  189;  Conservative  104;  Mismatches  261;  Indels   68;  Gaps   25;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWN---DPGAQYSLVDK--EKLPRFPFSIDQE-GDI 292
          |  | |:  :||   |:|::: |:: |   :     ||:  :  :  |   | |::| | :
Db    112 WVIP-PISCLENHRGPYPMRLVQIKSNKDKESKVYYSITGQGADSPPVGIFIIERETGWL 170

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
```

```
Db    171 EVTEQLDREKIDRYTLLSHAVSASGQPVEDPMEIIITVMDQNDNKPVFIKEVFVGYIEEN 230

Qy    353 ERLGNSIGTLTAHDRDEE-NTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLK 411

Db    231 AKPGTSVMTVNATDADDAVNTDNGIVSYSIVSQQPPRPHPQMFTIDPAKGIISVLGTGLD 290

Qy    412 KQDTPQYNLTIEVSDKDFKTL---CFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGST 468

Db    291 RETTPNYTLIVQATDQEGKGLSNTATAIIEVTDANDNIPIFNPTMYEG--VVEENKPGTE 348

Qy    469 ILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSN 528

Db    349 VARLTVTDQDAPGSPAWQAVYHIKSGNLDGAFSIITDPSTNNGILKTAKGLDYETKSRYD 408

Qy    529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588

Db    409 LVVTVENKVPLSVPI---TLSTASVLVTVLDVNEPPVFVPPIKRVGVPEDLPVGQQVTSY 465

Qy    589 TAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATE 641

Db    466 TAEDPDRDMRQKITYRMGSDPAGWLYIHPENGIVTATQPLDRESVHAINSTYKAIILAVD 525

Qy    642 VGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGP 701

Db    526 NGIPDTTGTGTLLLLLQDVNDNGP--TPEPRSFEICS--RQPEKQILSIVDKD----LPP 577

Qy    702 H---FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEG-- 756

Db    578 HTYPFKAALEHGS-SNNWTV-EIRGQDELAMGLKKELEPGEYNIFVKLTDS-----QGKA 630

Qy    757 -IVSLPVTFCSCVEG---SCFRPA----GHQTGIPTV-GMAVGILLTTLLVIGIILAVVF 807

Db    631 QVTQVKAQVCEC-EGTAKNCERRSYIVGG--LGVPAILGILGGIL---ALLILLLLLLLF 684

Qy    808 IRIKK----------DKGKDNV 819

Db    685 ARRRKVEKEPLLPPEDDMRDNV 706


RESULT 14
US-08-474-068A-7
; Sequence 7, Application US/08474068A
; Patent No. 5837525
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell & Flores LLP
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,068A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1683
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  837 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-474-068A-7

  Query Match             14.7%;  Score 634;  DB 2;  Length 837;
  Best Local Similarity   30.4%;  Pred. No. 2e-47;
  Matches  189;  Conservative  104;  Mismatches  261;  Indels   68;  Gaps   25;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWN---DPGAQYSLVDK--EKLPRFPFSIDQE-GDI 292

Db    112 WVIP-PISCLENHRGPYPMRLVQIKSNKDKESKVYYSITGQGADSPPVGIFIIERETGWL 170

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352

Db    171 EVTEQLDREKIDRYTLLSHAVSASGQPVEDPMEIIITVMDQNDNKPVFIKEVFVGYIEEN 230

Qy    353 ERLGNSIGTLTAHDRDEE-NTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLK 411

Db    231 AKPGTSVMTVNATDADDAVNTDNGIVSYSIVSQQPPRPHPQMFTIDPAKGIISVLGTGLD 290

Qy    412 KQDTPQYNLTIEVSDKDFKTL---CFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGST 468

Db    291 RETTPNYTLIVQATDQEGKGLSNTATAIIEVTDANDNIPIFNPTMYEG--VVEENKPGTE 348

Qy    469 ILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSN 528

Db    349 VARLTVTDQDAPGSPAWQAVYHIKSGNLDGAFSIITDPSTNNGILKTAKGLDYETKSRYD 408

Qy    529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588

Db    409 LVVTVENKVPLSVPI---TLSTASVLVTVLDVNEPPVFVPPIKRVGVPEDLPVGQQVTSY 465

Qy    589 TAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATE 641

Db    466 TAEDPDRDMRQKITYRMGSDPAGWLYIHPENGIVTATQPLDRESVHAINSTYKAIILAVD 525

Qy    642 VGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGP 701

Db    526 NGIPDTTGTGTLLLLLQDVNDNGP--TPEPRSFEICS--RQPEKQILSIVDKD----LPP 577

Qy    702 H---FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEG-- 756

Db    578 HTYPFKAALEHGS-SNNWTV-EIRGQDELAMGLKKELEPGEYNIFVKLTDS-----QGKA 630

Qy    757 -IVSLPVTFCSCVEG---SCFRPA----GHQTGIPTV-GMAVGILLTTLLVIGIILAVVF 807

Db    631 QVTQVKAQVCEC-EGTAKNCERRSYIVGG--LGVPAILGILGGIL---ALLILLLLLLLF 684

Qy    808 IRIKK----------DKGKDNV 819

Db    685 ARRRKVEKEPLLPPEDDMRDNV 706


RESULT 15
US-08-472-481-6
; Sequence 6, Application US/08472481
; Patent No. 5863804
;  GENERAL INFORMATION:
;    APPLICANT:  Ranscht, Barbara
;    TITLE OF INVENTION:  T-Cadherin Adhesion Molecule
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Campbell and Flores
;      STREET:  4370 La Jolla Village Drive, Suite 700
;      CITY:  San Diego
;      STATE:  California
```

```
;      COUNTRY:  United States
;      ZIP:  92122
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/472,481
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/213,361
;      FILING DATE:  14-MAY-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/607,293
;      FILING DATE:  30-OCT-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Campbell, Cathryn A.
;      REGISTRATION NUMBER:  31,815
;      REFERENCE/DOCKET NUMBER:  P-LJ 1686
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 535-9001
;      TELEFAX:  (619) 535-8949
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  837 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-472-481-6

  Query Match             14.7%;  Score 634;  DB 2;  Length 837;
  Best Local Similarity   30.4%;  Pred. No. 2e-47;
  Matches  189;  Conservative  104;  Mismatches  261;  Indels   68;  Gaps   25;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWN---DPGAQYSLVDK--EKLPRFPFSIDQE-GDI 292
          |  | |:  :||   |:|::: |:: |   :    ||:  :  :  |   | |::| | :
Db    112 WVIP-PISCLENHRGPYPMRLVQIKSNKDKESKVYYSITGQGADSPPVGIFIIERETGWL 170

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
           ||: |||| | |   : |    |:|:  |:|| : | | ||| |     | |  ::||
Db    171 EVTEQLDREKIDRYTLLSHAVSASGQPVEDPMEIIITVMDQNDNKPVFIKEVFVGYIEEN 230

Qy    353 ERLGNSIGTLTAHDRDEE-NTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLK 411
          : | |: |: | | |:  || |  ::| || | |  |   :| |    |:: :    |
Db    231 AKPGTSVMTVNATDADDAVNTDNGIVSYSIVSQQPPRPHPQMFTIDPAKGIISVLGTGLD 290

Qy    412 KQDTPQYNLTIEVSDKDFKTL---CFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGST 468
          :: || | | :: :|:: | |       | | | || ||||  : |    : |:   |:
Db    291 RETTPNYTLIVQATDQEGKGLSNTATAIIEVTDANDNIPIFNPTMYEG--VVEENKPGTE 348

Qy    469 ILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSN 528
          :  :  || | | : : : :|||  |: :|   : ||| || | :   | ||:|| :  :
Db    349 VARLTVTDQDAPGSPAWQAVYHIKSGNLDGAFSIITDPSTNNGILKTAKGLDYETKSRYD 408

Qy    529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
          :|   ||  ||   :     | |   : | |||| | |   : :  | ||: :| :| :
Db    409 LVVTVENKVPLSVPI---TLSTASVLVTVLDVNEPPVFVPPIKRVGVPEDLPVGQQVTSY 465

Qy    589 TAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATE 641
          ||:||:      |:| :  |  ||| |    | : :  |||||:     | |:  ::| :
Db    466 TAEDPDRDMRQKITYRMGSDPAGWLYIHPENGIVTATQPLDRESVHAINSTYKAIILAVD 525

Qy    642 VGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGP 701
           |    :      |:| |||||  |    :       |     |   |    | |       |
Db    526 NGIPDTTGTGTLLLLLQDVNDNGP--TPEPRSFEICS--RQPEKQILSIVDKD----LPP 577

Qy    702 H---FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEG-- 756
          |   |  :|  ||  |:| | :| |           : |   | : ::: |      :|
Db    578 HTYPFKAALEHGS-SNNWTV-EIRGQDELAMGLKKELEPGEYNIFVKLTDS-----QGKA 630

Qy    757 -IVSLPVTFCSCVEG---SCFRPA----GHQTGIPTV-GMAVGILLTTLLVIGIILAVVF 807
           :  :    | | ||   :| | :    |   |:| : |:  |||    |:| ::| ::|

Db    631 QVTQVKAQVCEC-EGTAKNCERRSYIVGG--LGVPAILGILGGIL---ALLILLLLLLLF 684

Qy    808 IRIKK----------DKGKDNV 819
            | :|          |  :|||
Db    685 ARRRKVEKEPLLPPEDDMRDNV 706


Search completed: January 15, 2003, 08:20:03
Job time : 30.6438 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:27 ; Search time 1.35616 Seconds
(without alignments)
954.610 Million cell updates/sec

Title: US-09-079-819-50
Perfect score: 249
Sequence: 1 EWYSWKRSSKSTGLGDTATR.....PSQSDGCPYNGRLTTVKPRK 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	55	22.1	5405	4	US-08-718-388-9	Sequence 9, Appli
2	54.5	21.9	412	1	US-08-313-288B-18	Sequence 18, Appl
3	53.5	21.5	263	1	US-07-971-096-4	Sequence 4, Appli
4	53.5	21.5	263	1	US-08-175-096-4	Sequence 4, Appli
5	53.5	21.5	263	4	US-08-413-974-6	Sequence 6, Appli
6	53.5	21.5	263	4	US-08-434-418-6	Sequence 6, Appli
7	53.5	21.5	263	4	US-08-433-288-6	Sequence 6, Appli
8	53.5	21.5	263	4	US-08-174-739A-6	Sequence 6, Appli
9	53.5	21.5	263	4	US-08-434-256-6	Sequence 6, Appli
10	53.5	21.5	596	4	US-09-199-637A-331	Sequence 331, App
11	53	21.3	1156	3	US-08-996-083-1	Sequence 1, Appli
12	53	21.3	1156	4	US-09-429-516-1	Sequence 1, Appli
13	53	21.3	1156	4	US-09-429-516-3	Sequence 3, Appli
14	52	20.9	553	4	US-09-249-697A-6	Sequence 6, Appli
15	52	20.9	553	4	US-09-249-697A-19	Sequence 19, Appl
16	52	20.9	553	4	US-09-363-316B-6	Sequence 6, Appli
17	52	20.9	553	4	US-09-363-316B-24	Sequence 24, Appl
18	51.5	20.7	788	2	US-08-918-914-4	Sequence 4, Appli
19	50.5	20.3	148	1	US-08-207-904-15	Sequence 15, Appl
20	50	20.1	43	2	US-08-488-161-70	Sequence 70, Appl
21	50	20.1	43	3	US-09-273-685-70	Sequence 70, Appl
22	50	20.1	43	5	PCT-US95-11934-70	Sequence 70, Appl
23	49.5	19.9	108	4	US-09-064-411A-34	Sequence 34, Appl
24	49.5	19.9	320	4	US-09-064-411A-44	Sequence 44, Appl
25	49.5	19.9	332	4	US-09-064-411A-39	Sequence 39, Appl
26	49.5	19.9	366	4	US-09-064-411A-37	Sequence 37, Appl
27	49.5	19.9	367	1	US-08-075-533-2	Sequence 2, Appli
28	49.5	19.9	367	1	US-08-075-533-4	Sequence 4, Appli
29	49.5	19.9	367	2	US-08-948-176-2	Sequence 2, Appli
30	49.5	19.9	367	2	US-08-948-176-4	Sequence 4, Appli
31	49.5	19.9	367	5	PCT-US91-09160-2	Sequence 2, Appli
32	49.5	19.9	367	5	PCT-US91-09160-4	Sequence 4, Appli
33	49	19.7	110	1	US-08-300-386A-70	Sequence 70, Appl
34	49	19.7	110	3	US-08-931-645-70	Sequence 70, Appl
35	49	19.7	110	5	PCT-US95-11235-70	Sequence 70, Appl
36	49	19.7	290	4	US-09-286-691-29	Sequence 29, Appl
37	49	19.7	290	4	US-09-687-147-29	Sequence 29, Appl
38	48.5	19.5	369	1	US-08-230-047-5	Sequence 5, Appli
39	48.5	19.5	455	3	US-09-038-217A-18	Sequence 18, Appl
40	48	19.3	33	4	US-08-469-260A-399	Sequence 399, App
41	48	19.3	50	1	US-07-903-456-6	Sequence 6, Appli
42	48	19.3	50	4	US-08-249-241-6	Sequence 6, Appli
43	48	19.3	137	1	US-08-392-828C-38	Sequence 38, Appl
44	48	19.3	137	3	US-09-330-945-38	Sequence 38, Appl
45	48	19.3	218	3	US-08-985-526-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
;  GENERAL INFORMATION:
;    APPLICANT: MORIKAWA, MINORU
;    APPLICANT: HARADA, NAOKI
;    TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING
;    TITLE OF INVENTION: PROTEIN
;    NUMBER OF SEQUENCES: 29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
;      STREET: PO BOX 747
;      CITY: FALLS CHURCH
;      STATE: VA
;      COUNTRY: USA
;      ZIP: 22040-0747
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/718,388
;      FILING DATE:
;      CLASSIFICATION: 536
;    ATTORNEY/AGENT INFORMATION:
;      NAME: MURPHY JR, GERALD M
;      REGISTRATION NUMBER: 28,977
;      REFERENCE/DOCKET NUMBER: 0230-111
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: (703) 205-8000
;      TELEFAX: (703) 205-8050
;  INFORMATION FOR SEQ ID NO: 9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 5405 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-718-388-9

  Query Match          22.1%;  Score 55;  DB 4;  Length 5405;
  Best Local Similarity  41.4%;  Pred. No. 2.9e+02;
  Matches   12;  Conservative    2;  Mismatches    7;  Indels    8;  Gaps    2;

Qy       4 SWKRSSKSTGLGDTATREGCGPSQSDGCP 32
           :|:    | | |     |||||    |||
Db    5014 AWRVPGSSKGCG-----EGCGP---QGCP 5034
```

```
RESULT 2
US-08-313-288B-18
; Sequence 18, Application US/08313288B
; Patent No. 5750502
;  GENERAL INFORMATION:
;    APPLICANT:  Jessell, Thomas M. and Avihu Klar
;    TITLE OF INVENTION:  CLONING, EXPRESSION AND USES OF A
;    TITLE OF INVENTION:  NOVEL SECRETED PROTEIN, F-SPONDIN
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/313,288B
;      FILING DATE:  January 5, 1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  40028-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0526
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  412 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-313-288B-18

  Query Match             21.9%;  Score 54.5;  DB 1;  Length 412;
  Best Local Similarity   36.1%;  Pred. No. 16;
  Matches   13;  Conservative    4;  Mismatches   12;  Indels    7;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTAT-------REGCGPSQSDG 30
           ||| |::|:| |  |           |||    : ||
Db      62 WYSLKKNSRSLGENDDGNNNNGDNGREGKDEDKRDG 97


RESULT 3
US-07-971-096-4
; Sequence 4, Application US/07971096
; Patent No. 5480972
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Knox, Robert Bruce
;    TITLE OF INVENTION:  ALLERGENIC PROTEINS AND PEPTIDES FROM
;    TITLE OF INVENTION:  JOHNSON GRASS POLLEN
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0,
;      SOFTWARE:  Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/971,096
;      FILING DATE:  19921030
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-042 (IMI-022)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  263 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-971-096-4

  Query Match             21.5%;  Score 53.5;  DB 1;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 12;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||       | || |       ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 76


RESULT 4
US-08-175-096-4
; Sequence 4, Application US/08175096
; Patent No. 5691167
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Knox, Robert Bruce
;    TITLE OF INVENTION:  ALLERGENIC PROTEINS AND PEPTIDES FROM
;    TITLE OF INVENTION:  JOHNSON GRASS POLLEN
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0,
;      SOFTWARE:  Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/175,096
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/971,096
;      FILING DATE:  OCT 30, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IPC-042 (IMI-022)
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  263 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-175-096-4

  Query Match             21.5%;  Score 53.5;  DB 1;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 12;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||       | || |       ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 76


RESULT 5
US-08-413-974-6
; Sequence 6, Application US/08413974
; Patent No. 6180368
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 6180368ris
;      STREET:  1 Liberty Place, 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/413,974
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/202,861
;      FILING DATE:
;      APPLICATION NUMBER:  US/07/746,703
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3949
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  263 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-413-974-6

  Query Match             21.5%;  Score 53.5;  DB 4;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 12;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||       | || |       ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 76


RESULT 6
US-08-434-418-6
; Sequence 6, Application US/08434418
; Patent No. 6197313
; GENERAL INFORMATION:
;  APPLICANT:  Singh, Mohan Bir et al.
;  TITLE OF INVENTION:  RYEGRASS POLLEN ALLERGEN
;  FILE REFERENCE:  IMI-051CND2
;  CURRENT APPLICATION NUMBER: US/08/434,418
;  CURRENT FILING DATE:  1995-05-03
;  PRIOR APPLICATION NUMBER:  08/202,861
;  PRIOR FILING DATE:  1994-25-02
;  NUMBER OF SEQ ID NOS:  25
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 263
;   TYPE: PRT
;   ORGANISM: Lolium perenne
US-08-434-418-6

  Query Match             21.5%;  Score 53.5;  DB 4;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 12;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||       | || |       ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 76


RESULT 7
US-08-433-288-6
; Sequence 6, Application US/08433288
; Patent No. 6239269
; GENERAL INFORMATION:
;  APPLICANT:  Singh, Mohan Bir et al.
;  TITLE OF INVENTION:  RYEGRASS POLLEN ALLERGEN
;  FILE REFERENCE:  IMI-051CND1
;  CURRENT APPLICATION NUMBER: US/08/433,288
;  CURRENT FILING DATE:  1995-05-03
;  PRIOR APPLICATION NUMBER:  08/413,947
;  PRIOR FILING DATE:  1995-03-30
;  PRIOR APPLICATION NUMBER:  08/202,861
;  PRIOR FILING DATE:  1994-02-25
;  PRIOR APPLICATION NUMBER:  07/746,703
;  PRIOR FILING DATE:  1991-08-16
;  PRIOR APPLICATION NUMBER:  07/585,086
;  PRIOR FILING DATE:  1990-10-26
;  NUMBER OF SEQ ID NOS:  25
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 263
;   TYPE: PRT
;   ORGANISM: Lolium perenne
US-08-433-288-6

  Query Match             21.5%;  Score 53.5;  DB 4;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 12;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||       | || |       ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 76


RESULT 8
US-08-174-739A-6
; Sequence 6, Application US/08174739A
; Patent No. 6265566
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir
;    APPLICANT:  Knox, Robert B.
;    APPLICANT:  Smith, Penelope
;    APPLICANT:  Avjioglu, Asil
```

```
;    APPLICANT:  Theerakulpisut, Piyada
;    APPLICANT:  Hough, Terryn
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lahive & Cockfield, LLP
;      STREET:  60 State Street, Suite 510
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/174,739A
;      FILING DATE:  29-DEC-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Mandragouras, Amy E.
;      REGISTRATION NUMBER:  36,207
;      REFERENCE/DOCKET NUMBER:  IMI-051DV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-7400
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  263 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-174-739A-6

  Query Match             21.5%;  Score 53.5;  DB 4;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 12;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||      | || |       ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 76


RESULT 9
US-08-434-256-6
; Sequence 6, Application US/08434256
; Patent No. 6451324
;  GENERAL INFORMATION:
;    APPLICANT:  Singh, Mohan Bir, Knox, Robert B., Smith, Penelope,
;    APPLICANT:  Avjioglu, Asil, Theerakulpisut, Piyada, Hough, Terryn
;    TITLE OF INVENTION:  Ryegrass Pollen Allergen
;    NUMBER OF SEQUENCES:  25
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz & No. 6451324ris
;      STREET:  1 Liberty Place, 46th Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/434,256
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hohenschutz, Liza D.
;      REGISTRATION NUMBER:  33,712
;      REFERENCE/DOCKET NUMBER:  IMPH-0003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215)568-3100
;      TELEFAX:  (215)568-3949
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  263 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-434-256-6

  Query Match             21.5%;  Score 53.5;  DB 4;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 12;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||      | || |       ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 76


RESULT 10
US-09-199-637A-331
; Sequence 331, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
;  APPLICANT:  Ausubel, Frederick
;  APPLICANT:  Goodman, Howard M.
;  APPLICANT:  Rahme, Laurence G.
;  APPLICANT:  Mahajan-Miklos, Shalina
;  APPLICANT:  Tan, Man-Wah
;  APPLICANT:  Cao, Hui
;  APPLICANT:  Drenkard, Eliana
;  APPLICANT:  Tsongalis, John
;  TITLE OF INVENTION:  VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE:  00786/361002
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE:  1997-11-25
;  NUMBER OF SEQ ID NOS:  437
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 331
;   LENGTH: 596
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-331

  Query Match             21.5%;  Score 53.5;  DB 4;  Length 596;
  Best Local Similarity   37.0%;  Pred. No. 33;
  Matches   17;  Conservative    4;  Mismatches   12;  Indels   13;  Gaps    2;

Qy       9 SKSTGLGDTATREGCGPSQS----------DGCPYNGRLTTVKPRK 44
           || ||||   :|| |  :|:           ||    |:|   |||
Db     552 SKGTGLGLYLSRELCESNQARIDYRNREEGGGC---FRITFAHPRK 594


RESULT 11
US-08-996-083-1
; Sequence 1, Application US/08996083A
; Patent No. 6124095
; GENERAL INFORMATION:
;  APPLICANT: Magna, Holly
;  APPLICANT: Schaffer, Paul
;  APPLICANT: Lawton, Michael
;  APPLICANT: Yocum, Sue
;  APPLICANT: Mitchell, Peter
;  APPLICANT: Hutchinson, Nancy
;  APPLICANT: Murry, Lynn E.
;  TITLE OF INVENTION: HUMAN NUCLEOTIDE PYROPHOSPHOHYDROLASE-2
;  FILE REFERENCE: PF-0420 US
;  CURRENT APPLICATION NUMBER: US/08/996,083A
;  CURRENT FILING DATE: 1997-12-22
```

```
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1156
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte Clone No. 6124095: 1388013
;  PUBLICATION INFORMATION:
US-08-996-083-1

  Query Match             21.3%;  Score 53;  DB 3;  Length 1156;
  Best Local Similarity   34.8%;  Pred. No. 86;
  Matches   16;  Conservative    1;  Mismatches   23;  Indels    6;  Gaps    2;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGR----LTTVKPR 43
           |  |   | | | |    |  |     | ||  ||       |:||
Db     152 WGPWGPCSGSCGPGRRLRRRHCPSPAGDACP--GRPLEAQKCVRPR 195


RESULT 12
US-09-429-516-1
; Sequence 1, Application US/09429516
; Patent No. 6251389
;  GENERAL INFORMATION:
;    APPLICANT:  Magna, Holly
;    APPLICANT:  Schaffer, Paul
;    APPLICANT:  Lawton, Michael
;    APPLICANT:  Yocum, Sue
;    APPLICANT:  Mitchell, Peter
;    APPLICANT:  Hutchinson, Nancy
;    APPLICANT:  Murry, Lynn E.
;    TITLE OF INVENTION:  HUMAN NUCLEOTIDE PYROPHOS
;    TITLE OF INVENTION:  PHOHYDROLASE-2
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/429,516
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/996,083
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0420 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1156 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  SATPF1008
;      CLONE:  1388013
US-09-429-516-1

  Query Match             21.3%;  Score 53;  DB 4;  Length 1156;
  Best Local Similarity   34.8%;  Pred. No. 86;
  Matches   16;  Conservative    1;  Mismatches   23;  Indels    6;  Gaps    2;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGR----LTTVKPR 43
           |  |   | | | |    |  |     | ||  ||       |:||
Db     152 WGPWGPCSGSCGPGRRLRRRHCPSPAGDACP--GRPLEAQKCVRPR 195


RESULT 13
US-09-429-516-3
; Sequence 3, Application US/09429516
; Patent No. 6251389
;  GENERAL INFORMATION:
;    APPLICANT:  Magna, Holly
;    APPLICANT:  Schaffer, Paul
;    APPLICANT:  Lawton, Michael
;    APPLICANT:  Yocum, Sue
;    APPLICANT:  Mitchell, Peter
;    APPLICANT:  Hutchinson, Nancy
;    APPLICANT:  Murry, Lynn E.
;    TITLE OF INVENTION:  HUMAN NUCLEOTIDE PYROPHOS
;    TITLE OF INVENTION:  PHOHYDROLASE-2
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/429,516
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/996,083
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0420 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1156 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      LIBRARY:  SATPF1002
;      CLONE:  422069
US-09-429-516-3

  Query Match             21.3%;  Score 53;  DB 4;  Length 1156;
  Best Local Similarity   34.8%;  Pred. No. 86;
  Matches   16;  Conservative    1;  Mismatches   23;  Indels    6;  Gaps    2;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGR----LTTVKPR 43
           |  |   | | | |    |  |     | ||  ||       |:||
Db     152 WGPWGPCSGSCGPGRRLRRRHCPSPAGDACP--GRPLEAQKCVRPR 195


RESULT 14
```

```
US-09-249-697A-6
; Sequence 6, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
;  TITLE OF INVENTION:  LIVER SPLEEN
;  FILE REFERENCE: 24011-727
;  CURRENT APPLICATION NUMBER: US/09/249,697A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(553)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-249-697A-6

  Query Match             20.9%;  Score 52;  DB 4;  Length 553;
  Best Local Similarity   29.1%;  Pred. No. 49;
  Matches   16;  Conservative    9;  Mismatches   14;  Indels   16;  Gaps    4;

Qy       3 YSWKRSSKSTGLGDTATREGC------GPSQSDGCP-YNGRLTT-------VKPR 43
           | |:|:||  |: :     ||      ||::    | | |:  :       :|||
Db      51 YGWRRNSK--GVCEATCEPGCKFGECVGPNKCRCFPGYTGKTCSQDVNECGMKPR 103


RESULT 15
US-09-249-697A-19
; Sequence 19, Application US/09249697A
; Patent No. 6392018
; GENERAL INFORMATION:
;  APPLICANT: Ford, John
;  APPLICANT:  Yeung, George
;  TITLE OF INVENTION: NOVEL EGF MOTIF PROTEIN OBTAINED FROM A CDNA LIBRARY OF FETAL
;  TITLE OF INVENTION:  LIVER SPLEEN
;  FILE REFERENCE: 24011-727
;  CURRENT APPLICATION NUMBER: US/09/249,697A
;  CURRENT FILING DATE:  1999-02-12
;  PRIOR APPLICATION NUMBER: US 08/968,800
;  PRIOR FILING DATE: 1997-11-22
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 19
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-249-697A-19

  Query Match             20.9%;  Score 52;  DB 4;  Length 553;
  Best Local Similarity   29.1%;  Pred. No. 49;
  Matches   16;  Conservative    9;  Mismatches   14;  Indels   16;  Gaps    4;

Qy       3 YSWKRSSKSTGLGDTATREGC------GPSQSDGCP-YNGRLTT-------VKPR 43
           | |:|:||  |: :     ||      ||::    | | |:  :       :|||
Db      51 YGWRRNSK--GVCEATCEPGCKFGECVGPNKCRCFPGYTGKTCSQDVNECGMKPR 103


Search completed: January 15, 2003, 08:19:58
Job time : 3.35616 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:24 ; Search time 3.31507 Seconds
(without alignments)
1768.599 Million cell updates/sec

Title: US-09-079-819-50
Perfect score: 249
Sequence: 1 EWYSWKRSSKSTGLGDTATR.....PSQSDGCPYNGRLTTVKPRK 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	249	100.0	44	20	AAW74068	Human HPT-1 bindin
2	200	80.3	44	20	AAW74045	Human D2H binding
3	68	27.3	793	22	ABB58866	Drosophila melanog
4	65	26.1	206	21	AAB08135	Amino acid sequenc
5	65	26.1	1522	20	AAW99302	Human BAI3 protein
6	65	26.1	1522	21	AAB23601	Human secreted pro
7	64.5	25.9	1572	20	AAW99301	Human BAI2 protein
8	59	23.7	275	22	ABB63063	Drosophila melanog
9	58.5	23.5	184	11	AAR07289	Circumsporozoite a
10	58.5	23.5	710	22	ABG20114	Novel human diagno
11	58.5	23.5	1068	22	AAG67244	Amino acid sequenc
12	58	23.3	226	20	AAW86332	Kidney injury asso
13	58	23.3	244	22	ABG05161	Novel human diagno
14	57.5	23.1	120	22	ABG16825	Novel human diagno
15	57.5	23.1	183	23	AAU74796	Mouse thrombospond
16	57.5	23.1	1172	23	AAU74786	Mouse thrombospond
17	57	22.9	70	22	AAO03670	Human polypeptide
18	57	22.9	469	21	AAB43834	Human cancer assoc
19	57	22.9	469	22	AAG75603	Human colon cancer
20	57	22.9	534	22	ABG14674	Novel human diagno
21	57	22.9	735	22	ABG22215	Novel human diagno
22	57	22.9	1092	23	ABG34077	Human Pro peptide
23	56.5	22.7	208	23	AAB47770	Human thrombospond
24	56.5	22.7	1093	22	AAE02455	Mouse semaphorin G
25	56.5	22.7	1336	23	AAB47771	Human thrombospond
26	56.5	22.7	5635	23	ABP60991	Novel human protei
27	56	22.5	1705	22	ABB65443	Drosophila melanog
28	56	22.5	7337	22	ABG22216	Novel human diagno
29	55.5	22.3	939	23	AAG68296	Human semaphorin G
30	55.5	22.3	954	23	AAG68295	Human semaphorin G
31	55.5	22.3	999	21	AAY94990	Human secreted pro
32	55.5	22.3	1034	23	AAG68291	Human semaphorin G
33	55.5	22.3	1049	23	AAG68289	Human semaphorin G
34	55.5	22.3	1078	23	AAG68292	Human semaphorin G
35	55.5	22.3	1093	23	AAG68290	Human semaphorin G
36	55.5	22.3	1136	23	AAG68294	Human semaphorin G
37	55.5	22.3	1151	23	AAG68293	Human semaphorin G
38	55.5	22.3	1352	23	AAE18212	Human MOL4 protein
39	55	22.1	55	23	AAB47773	Human thrombospond
40	55	22.1	119	22	AAM94580	Human reproductive
41	55	22.1	208	22	AAO07478	Human polypeptide
42	55	22.1	482	22	AAG73853	Human colon cancer
43	55	22.1	5405	16	AAW14749	IgG-Fc binding pro
44	54.5	21.9	309	12	AAR13175	NS1_81-RLfdelta9.
45	54.5	21.9	319	11	AAR07945	NS181RLFAuth plasm

ALIGNMENTS

RESULT 1
AAW74068
ID AAW74068 standard; peptide; 44 AA.
XX
AC AAW74068;
XX
DT 04-MAY-1999 (first entry)
XX
DE Human HPT-1 binding protein HAX35.
XX
KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy.
XX
OS Homo sapiens.
XX
PN WO9851325-A2.
XX
PD 19-NOV-1998.
XX
PF 15-MAY-1998; 98WO-US10088.
XX
PR 15-MAY-1997; 97US-0046595.
XX
PA (CYTO-) CYTOGEN CORP.
PA (ELAN-) ELAN CORP PLC.
XX
PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
PI Omahony DJ, Patterson CA, Singleton J;
XX
DR WPI; 1999-009568/01.

```
XX
PT   New proteins that bind specifically to receptors in the
PT   gastro-intestinal tract and related nucleic acid - chimaeras and
PT   antibodies, used to deliver therapeutic or diagnostic agents to, or
PT   through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
PS   Claim 2; Page 55; 294pp; English.
XX
CC   This sequence represents a peptide that specifically binds to the human
CC   HPT-1 protein. The invention relates to purified proteins (I) that bind
CC   specifically to at least one of the gastro-intestinal (GI) tract
CC   receptors human intestinal peptide-associated transporter (HPT1),
CC   hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
CC   active transport of therapeutic agents through human and animal GI tissue
CC   (into the blood) for in vivo delivery, particularly for treatment or
CC   prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
CC   cancer, migraine, or angina pectoris. Specifically they are used to
CC   deliver insulin or leuprolide, but many other suitable therapeutic agents
CC   are disclosed, including genes or inhibitory nucleic acid, imaging agents
CC   and antigens. (I) may also provide targeting to the GI tract. Other uses
CC   of (I) are: (i) to determine the level of specified receptors in a sample
CC   (in a binding assay); and (ii) to screen for molecules that bind (I).
CC   Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC   and in immunoassays. The antibodies are used to locate, detect and
CC   measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC   etc., also for peptide purification and immobilisation.
XX
SQ   Sequence   44 AA;

  Query Match             100.0%;  Score 249;  DB 20;  Length 44;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-26;
  Matches   44;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPRK 44
          |||||||||||||||||||||||||||||||||||||||||||||
Db      1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPRK 44


RESULT 2
AAW74045
ID   AAW74045 standard; peptide; 44 AA.
XX
AC   AAW74045;
XX
DT   04-MAY-1999  (first entry)
XX
DE   Human D2H binding protein DCX36.
XX
KW   Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW   D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW   intestinal peptide-associated transporter; hypertension; diabetes;
KW   osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW   therapeutic agent delivery; therapy.
XX
OS   Homo sapiens.
XX
PN   WO9851325-A2.
XX
PD   19-NOV-1998.
XX
PF   15-MAY-1998;   98WO-US10088.
XX
PR   15-MAY-1997;   97US-0046595.
XX
PA   (CYTO-) CYTOGEN CORP.
PA   (ELAN-) ELAN CORP PLC.
XX
PI   Alvarez VL,  Belinka BA,  Cagney GM,  Carter JM,  Lambkin IJ;
PI   Omahony DJ,  Patterson CA,  Singleton J;
XX
DR   WPI; 1999-009568/01.
XX
PT   New proteins that bind specifically to receptors in the
PT   gastro-intestinal tract and related nucleic acid - chimaeras and
PT   antibodies, used to deliver therapeutic or diagnostic agents to, or
PT   through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
PS   Claim 2; Page 55; 294pp; English.
XX
CC   This sequence represents a peptide that specifically binds to the human
CC   D2H protein. The invention relates to purified proteins (I) that bind
CC   specifically to at least one of the gastro-intestinal (GI) tract
CC   receptors human intestinal peptide-associated transporter (HPT1),
CC   hPEPT1, D2H and human sucrose-isomaltase complex (hSI). (I) provide
CC   active transport of therapeutic agents through human and animal GI tissue
CC   (into the blood) for in vivo delivery, particularly for treatment or
CC   prevention of hypertension, diabetes, osteoporosis, haemophilia, anaemia,
CC   cancer, migraine, or angina pectoris. Specifically they are used to
CC   deliver insulin or leuprolide, but many other suitable therapeutic agents
CC   are disclosed, including genes or inhibitory nucleic acid, imaging agents
CC   and antigens. (I) may also provide targeting to the GI tract. Other uses
CC   of (I) are: (i) to determine the level of specified receptors in a sample
CC   (in a binding assay); and (ii) to screen for molecules that bind (I).
CC   Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC   and in immunoassays. The antibodies are used to locate, detect and
CC   measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC   etc., also for peptide purification and immobilisation.
XX
SQ   Sequence   44 AA;

  Query Match             80.3%;  Score 200;  DB 20;  Length 44;
  Best Local Similarity   85.7%;  Pred. No. 5.8e-20;
  Matches   36;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPR 43
          | ||  |||:| ||| |||||||||||||||||||||||||
Db      2 WSSWNWSSKTTRLGDRATREGCGPSQSDGCPYNGRLTTVKPR 43


RESULT 3
ABB58866
ID   ABB58866 standard; Protein; 793 AA.
XX
AC   ABB58866;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 3390.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02969.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 3390; 21pp + Sequence Listing; English.
XX
```

```
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   793 AA;

  Query Match             27.3%;  Score 68;  DB 22;  Length 793;
  Best Local Similarity   26.3%;  Pred. No. 2.1;
  Matches   10;  Conservative   10;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLT 38
             :|:  :: |   |    :: : | |:  :|| || : |
Db     400 KWHKKRKESLKLGFARFSSSDDCAPAYGEGCAYNWKQT 437


RESULT 4
AAB08135
ID   AAB08135 standard; Protein; 206 AA.
XX
AC   AAB08135;
XX
DT   04-DEC-2000  (first entry)
XX
DE   Amino acid sequence of a KIAA0550 polypeptide.
XX
KW   Type 1 repeat; thrombospondin-1; TSP-1; Type I repeat peptide; TRP;
KW   KIAA0688; KIAA0550; angiogenesis inhibitory protein; angiogenesis;
KW   cancer; tumour; rheumatoid arthritis; psoriasis; retinopathy;
KW   ocular angiogenic disease; macular degeneration; corneal graft rejection;
KW   neovascular glaucoma; retrolental fibroplasia; rubeosis;
KW   Osler-Webber Syndrome; myocardial angiogenesis; haemophiliac joint;
KW   plaque neovascularisation; telangiectasia; angiofibroma;
KW   wound granulation; coronary collateral; cerebral collateral;
KW   arteriovenous malformation; ischemic limb angiogenesis;
KW   neovascular glaucoma; retrolental fibroplasia; fracture; vasculogenesis;
KW   heliobacter related disease; hematopoiesis; ovulation; menstruation;
KW   placentation; birth control; embryo implantation.
XX
OS   Homo sapiens.
XX
PN   WO200047622-A2.
XX
PD   17-AUG-2000.
XX
PF   15-FEB-2000; 2000WO-GB00520.
XX
PR   15-FEB-1999;   99GB-0003408.
XX
PA   (OXFO-) OXFORD BIOMEDICA UK LTD.
XX
PI   Mazarakis N,  Martin-Rendon E,  Kingsman SM;
XX
DR   WPI; 2000-549137/50.
XX
PT   Non-naturally occurring Type I repeat peptide (TRP) derived from human
PT   thrombospondin-1, KIAA0688 or KIAA0550 proteins, useful in the
PT   treatment of angiogenesis and/or cancer -
XX
PS   Claim 5; Fig 7; 84pp; English.
XX
CC   The present sequence represents a KIAA0550 polypeptide. The
CC   specification describes a non-naturally occurring Type I repeat peptide
CC   (TRP), derived from human thrombospondin-1 (hTSP-1), KIAA0688 or KIAA0550
CC   proteins. The TRP protein is an angiogenesis inhibitory protein. TRP
CC   proteins are used for the treatment of conditions associated with
CC   angiogenesis and cancer. Angiogenic mediated diseases include tumours,
CC   rheumatoid arthritis, psoriasis, ocular angiogenic diseases, retinopathy
CC   of prematurity, macular degeneration, corneal graft rejection,
CC   neovascular glaucoma, retrolental fibroplasia, rubeosis), Osler-Webber
CC   Syndrome, myocardial angiogenesis, plaque neovascularisation,
CC   telangiectasia, haemophiliac joints, angiofibroma, wound granulation,
CC   coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC   ischemic limb angiogenesis, neovascular glaucoma, retrolental
CC   fibroplasia, heliobacter related diseases, fractures, vasculogenesis,
CC   hematopoiesis, ovulation, menstruation and placentation. TRPs are also
CC   useful in the treatment of disease of excessive or abnormal stimulation
CC   of endothelial cells. TRP can also be used as a birth control agent, as
CC   it prevents the vascularisation required for embryo implantation.
XX
SQ   Sequence   206 AA;

  Query Match             26.1%;  Score 65;  DB 21;  Length 206;
  Best Local Similarity   35.7%;  Pred. No. 1.1;
  Matches   15;  Conservative    1;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKP 42
           ||  |   | : | |       | | |  | |  |  |  ||
Db      55 EWSPWSLCSFTCGRGQRTRTRSCTPPQYGGRPCEGPETHHKP 96


RESULT 5
AAW99302
ID   AAW99302 standard; Protein; 1522 AA.
XX
AC   AAW99302;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human BAI3 protein.
XX
KW   Human; BAI3; brain; cancer; drug; diagnosis; prevention; treatment.
XX
OS   Homo sapiens.
XX
PN   JP11032766-A.
XX
PD   09-FEB-1999.
XX
PF   16-JUN-1997;   97JP-0176485.
XX
PR   23-MAY-1997;   97JP-0150460.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
DR   WPI; 1999-183823/16.
DR   N-PSDB; AAX21357.
XX
PT   New human BAI gene  - is expressed in brain plays important role in
PT   cancer formation
XX
PS   Claim 2; Page 26-30; 62pp; Japanese.
XX
CC   This sequence represents the human BAI3 protein.  The gene is expressed
CC   specifically in the brain and plays an important role in cancer formation
CC   in the brain.  The BAI3 protein can be used in drug compositions to
CC   diagnose, prevent or treat such cancers.
XX
SQ   Sequence   1522 AA;

  Query Match             26.1%;  Score 65;  DB 20;  Length 1522;
  Best Local Similarity   35.7%;  Pred. No. 12;
  Matches   15;  Conservative    1;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKP 42
           ||  |   | : | |       | | |  | |  |  |  ||
Db     350 EWSPWSLCSFTCGRGQRTRTRSCTPPQYGGRPCEGPETHHKP 391
```

```
RESULT 6
AAB23601
ID   AAB23601 standard; Protein; 1522 AA.
XX
AC   AAB23601;
XX
DT   12-JAN-2001  (first entry)
XX
DE   Human secreted protein SEQ ID NO: 2.
XX
KW   Human; secreted protein; cytokine; cell proliferation;
KW   nutritional supplement; immune modulation; autoimmune disorder;
KW   haematopoiesis regulation; tissue growth; haemostasis; inflammation.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         3..15
FT                   /label= signal_peptide
FT   Protein         16..1522
FT                   /label= mature_protein
XX
PN   WO200049134-A1.
XX
PD   24-AUG-2000.
XX
PF   18-FEB-2000; 2000WO-US04340.
XX
PR   19-FEB-1999;   99US-0120680.
PR   23-APR-1999;   99US-0298733.
PR   17-AUG-1999;   99US-0149639.
PR   23-SEP-1999;   99US-0155686.
PR   01-OCT-1999;   99US-0157247.
PR   29-NOV-1999;   99US-0167822.
PR   29-NOV-1999;   99US-0167823.
PR   15-FEB-2000; 2000US-0298733.
XX
PA   (ALPH-) ALPHAGENE INC.
XX
PI   Valenzuela D,  Yuan O,  Hoffman H,  Hall J,  Rapiejko P;
XX
DR   WPI; 2000-549267/50.
DR   N-PSDB; AAA93101.
XX
PT   New secreted proteins and polynucleotides encoding them, which are
PT   derived from Homo sapiens, useful for therapy, diagnosis, and research,
PT   as well as nutritional sources or supplements -
XX
PS   Claim 9; Page 234-238; 309pp; English.
XX
CC   The present sequence is the sequence of a human secreted protein. Its
CC   cDNA was isolated from a foetal brain cDNA library. The proteins
CC   and coding sequences of the invention can be used in the isolation of
CC   similar genes and proteins, in the elucidation of their function in vivo,
CC   and to treat a number of conditions. It is possible that they may have
CC   uses as nutritional supplements, as cytokine or cell proliferation
CC   factors, in immune modulation, where they may be used to treat immune and
CC   autoimmune diseases, as haematopoiesis regulators (treating myeloid or
CC   lymphoid cell deficiencies), in the promotion of tissue growth, they may
CC   have chemokine or chemotactic activity, haemostatic or thrombolytic
CC   activity, or anti-inflammatory activity.
XX
SQ   Sequence   1522 AA;

  Query Match             26.1%;  Score 65;  DB 21;  Length 1522;
  Best Local Similarity   35.7%;  Pred. No. 12;
  Matches   15;  Conservative    1;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKP 42
           ||  |   | : | |       | | |  | |  |  |  ||
Db     350 EWSPWSLCSFTCGRGQRTRTRSCTPPQYGGRPCEGPETHHKP 391


RESULT 7
AAW99301
ID   AAW99301 standard; Protein; 1572 AA.
XX
AC   AAW99301;
XX
DT   21-MAY-1999  (first entry)
XX
DE   Human BAI2 protein.
XX
KW   Human; BAI2; brain; cancer; drug; diagnosis; prevention; treatment.
XX
OS   Homo sapiens.
XX
PN   JP11032766-A.
XX
PD   09-FEB-1999.
XX
PF   16-JUN-1997;   97JP-0176485.
XX
PR   23-MAY-1997;   97JP-0150460.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
DR   WPI; 1999-183823/16.
DR   N-PSDB; AAX21356.
XX
PT   New human BAI gene  - is expressed in brain plays important role in
PT   cancer formation
XX
PS   Claim 2; Page 22-26; 62pp; Japanese.
XX
CC   This sequence represents the human BAI2 protein.  The gene is expressed
CC   specifically in the brain and plays an important role in cancer formation
CC   in the brain.  The BAI2 protein can be used in drug compositions to
CC   diagnose, prevent or treat such cancers.
XX
SQ   Sequence   1572 AA;

  Query Match             25.9%;  Score 64.5;  DB 20;  Length 1572;
  Best Local Similarity   28.3%;  Pred. No. 15;
  Matches   15;  Conservative    3;  Mismatches   18;  Indels   17;  Gaps    1;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDG----------------CPYNGR 36
           || ||   |:| | |  :    | | |  |                 ||  |:
Db     357 EWGSWSLCSRSCGRGSRSRMRTCVPPQHGGKACEGPELQTKLCSMAACPVEGQ 409


RESULT 8
ABB63063
ID   ABB63063 standard; Protein; 275 AA.
XX
AC   ABB63063;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 15981.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
```

XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL07166.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 15981; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 275 AA;

Query Match 23.7%; Score 59; DB 22; Length 275;
Best Local Similarity 40.7%; Pred. No. 10;
Matches 11; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 3 YSWKRSSKSTGLGDTATREGCGPSQSD 29
| ||| :|: | : | | ||: |
Db 209 YPWKRHNKNRGNYEPELRYACTPSKED 235

RESULT 9
AAR07289
ID AAR07289 standard; protein; 184 AA.
XX
AC AAR07289;
XX
DT 28-JAN-1991 (first entry)
XX
DE Circumsporozoite analogue Falciparum 3.
XX
KW CS protein; plasmodium; malaria; vaccine.
XX
OS Synthetic.
XX
PN EP392820-A.
XX
PD 17-OCT-1990.
XX
PF 11-APR-1990; 90EP-0303907.
XX
PR 11-APR-1989; 89US-0336288.
XX
PA (CHIR-) CHIRON CORP.
XX
PI Barr PJ, Bathurst IC, Gibson HL;
XX
DR WPI; 1990-314486/42.
DR N-PSDB; AAQ06152.
XX
PT Recombinant plasmodium circumsporozoite analogues - lacking
PT one or more repeat epitope(s) for use as a malaria sub-unit
PT vaccine.
XX
PS Claim 10; Fig 8; 22pp; English.
XX
CC The protein was produced by expression of a synthetic gene. The
CC analogue comprises AAs 43-123 of the native P. falciparum CS
CC protein, followed by four repeat sequences (three "B"s, i.e. NANP)
CC and one "A", i.e. NVDP), followed by AAs 289-374 of the native
CC protein. Reduction of the immunological dominance of the repeats
CC relative to the epitopes in the regions flanking the repeats
CC enhances sporozoite neutralising activity.
CC See also AAR07287-91.
XX
SQ Sequence 184 AA;

Query Match 23.5%; Score 58.5; DB 11; Length 184;
Best Local Similarity 38.9%; Pred. No. 7.4;
Matches 14; Conservative 4; Mismatches 11; Indels 7; Gaps 1;

Qy 2 WYSWKRSSKSTGLGDTAT-------REGCGPSQSDG 30
||| |::|:| | | | ||| : ||
Db 21 WYSLKKNSRSLGENDDANNNNGDNGREGKDEDKRDG 56

RESULT 10
ABG20114
ID ABG20114 standard; Protein; 710 AA.
XX
AC ABG20114;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20105.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS84301.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 50473; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

```
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   710 AA;

  Query Match             23.5%;  Score 58.5;  DB 22;  Length 710;
  Best Local Similarity   33.3%;  Pred. No. 38;
  Matches   13;  Conservative    4;  Mismatches   13;  Indels    9;  Gaps    1;

Qy      2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTV 40
          |:||::      || :||  ||           || | |
Db      1 WHSWRQQGWGAHLGSSATGPGC---------XQGRPTCV 30


RESULT 11
AAG67244
ID   AAG67244 standard; Protein; 1068 AA.
XX
AC   AAG67244;
XX
DT   13-NOV-2001  (first entry)
XX
DE   Amino acid sequence of murine thrombospondin 1-like protein.
XX
KW   Thrombospondin 1-like protein; TSP1-like protein; vesicle transport;
KW   brain tumour.
XX
OS   Mus musculus.
XX
PN   WO200109321-A1.
XX
PD   08-FEB-2001.
XX
PF   28-JUL-2000; 2000WO-JP05068.
XX
PR   29-JUL-1999;   99JP-0248036.
PR   27-AUG-1999;   99JP-0300253.
PR   18-OCT-1999;   99US-0159590.
PR   11-JAN-2000; 2000JP-0118776.
PR   17-FEB-2000; 2000US-0183322.
PR   02-MAY-2000; 2000JP-0183767.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T,  Murakami K;
PI   Yano K,  Kanzaki K,  Inoue Y;
XX
DR   WPI; 2001-541222/60.
DR   N-PSDB; AAH77754.
XX
PT   Gene encoding thrombospondin-like protein, and the protein and
PT   antibodies to it, useful for diagnosis and treatment of brain tumours -
XX
PS   Claim 2; Page 72-80; 105pp; Japanese.
XX
CC   The present sequence represents a thrombospondin 1-like (TSP1-like)
CC   protein. The cDNA sequence encoding human TSP1-like protein was isolated
CC   from a human 10 week-aged foetal tissue cDNA library. The TSP1-like
CC   protein may be involved in intracellular vesicle transport. Secretion
CC   of the TSP1-like protein is reduced in brain tumours. It can be used
CC   in the screening of target compounds, and is useful for diagnosis,
CC   prediction and treatment of brain tumour.
XX
SQ   Sequence   1068 AA;

  Query Match             23.5%;  Score 58.5;  DB 22;  Length 1068;
  Best Local Similarity   28.6%;  Pred. No. 61;
  Matches   14;  Conservative    7;  Mismatches   21;  Indels    7;  Gaps    2;

Qy      1 EWYSWKRSSKSTGLGDTATREGCG--PSQSDGCPYNGR-----LTTVKP 42
          ||  |:  |:| | |:      |   |:|  | |  |      :  ::|
Db    141 EWSFWEDCSRSCGHGNQTRTRTCSNPPAQHGGRPCEGHAVETIMCNIRP 189


RESULT 12
AAW86332
ID   AAW86332 standard; Protein; 226 AA.
XX
AC   AAW86332;
XX
DT   01-MAR-1999  (first entry)
XX
DE   Kidney injury associated molecule HW088 protein.
XX
KW   Kidney injury associated molecule; kidney injury related molecule;
KW   KIM; tissue growth promotion; regeneration; renal condition;
KW   acute renal failure; acute nephritis; tumour.
XX
OS   Rattus sp.
XX
PN   WO9853071-A1.
XX
PD   26-NOV-1998.
XX
PF   22-MAY-1998;   98WO-US10547.
XX
PR   23-MAY-1997;   97US-0047491.
PR   23-MAY-1997;   97US-0047490.
XX
PA   (BIOJ ) BIOGEN INC.
XX
PI   Cate RL,  Hession CA,  Sanicola-Nadel M,  Wei H;
XX
DR   WPI; 1999-045312/04.
DR   N-PSDB; AAV80626.
XX
PT   Kidney injury-associated molecule, KIM, polypeptides - upregulated
PT   in injured or regenerating tissues, useful to promote tissue growth
PT   and regeneration, especially to treat renal conditions
XX
PS   Claim 17; Page 168; 213pp; English.
XX
CC   The present sequence represents a kidney injury associated molecule
CC   (KIM) protein. KIM proteins can be administered therapeutically
CC   by expressing KIM encoding polynucleotides, to promote growth and/or
CC   survival of damaged tissue (e.g. renal tissue), since the KIM proteins
CC   are upregulated in injured or regenerating (especially renal) tissues.
CC   KIM fusion proteins, conjugates, antibodies and vectors can also be used
CC   therapeutically, e.g. these or the KIM proteins may be included with an
CC   acceptable carrier in pharmaceutical compositions, useful for therapy/
CC   prophylaxis of conditions associated with disfunction/disregulation of
CC   KIM genes or proteins, especially renal diseases or impairments of renal
CC   function in humans (e.g. acute renal failure, acute nephritis). The
CC   polynucleotides can be used to produce antisense sequences which, when
CC   internalised into cells, can disrupt expression of a cellular KIM gene,
CC   also useful in therapy (e.g. to block the growth of tumours dependent on
CC   KIM for growth) or compositions. The proteins and polynucleotides are
CC   useful diagnostically e.g. to detect and quantify renal injury/disease
CC   (indicative of increased risk, or presence of, renal injury or impaired
CC   function), or abnormal responses to tissue injury (indicative of
CC   increased risk, or presence of, an autoimmune response or abnormal
CC   tissue growth arising from/affecting renal tissue). The proteins can
CC   also be used to locate KIM-producing cells (especially specific loci,
CC   e.g. tissue masses abnormally producing/expressing KIM such as tumours
CC   arising from/affecting renal tissue), by contacting cells with an
CC   imageable KIM-binding reagent and imaging reagent accumulation.
XX
SQ   Sequence   226 AA;

  Query Match             23.3%;  Score 58;  DB 20;  Length 226;
  Best Local Similarity   40.9%;  Pred. No. 11;
  Matches   18;  Conservative    3;  Mismatches   17;  Indels    6;  Gaps    2;

Qy      7 RSSKSTGLGDTA-TREGC-----GPSQSDGCPYNGRLTTVKPRK 44
```

```
          ||   | |   | | : |     || ||||  |  |  ::|| |
Db      13 RSGLRTALAPRALTPQVCSPFATGPRQSDGTLYEFRTYSLKPSK 56
```

RESULT 13
ABG05161
ID ABG05161 standard; Protein; 244 AA.
XX
AC ABG05161;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #5152.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS69348.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 35520; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 244 AA;

```
  Query Match           23.3%;  Score 58;  DB 22;  Length 244;
  Best Local Similarity   34.1%;  Pred. No. 12;
  Matches   14;  Conservative    4;  Mismatches   17;  Indels    6;  Gaps    1;

Qy       4 SWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPRK 44
           ||: :|   |        |||  | ||    ||    :||:
Db     133 SWQGTSCGGG------TSGCGRLQXDGVASRGRWDPARPRE 167
```

RESULT 14
ABG16825
ID ABG16825 standard; Protein; 120 AA.
XX
AC ABG16825;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #16816.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS81012.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 47184; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 120 AA;

```
  Query Match           23.1%;  Score 57.5;  DB 22;  Length 120;
  Best Local Similarity   38.1%;  Pred. No. 6.1;
  Matches   16;  Conservative    5;  Mismatches   18;  Indels    3;  Gaps    2;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPR 43
           | |    :  ||   :||: || |  |  || : |  :| ||
Db      19 WRSPCCPAPKTGAWASATQGGC-PGPSSSCPASSR--SVSPR 57
```

```
RESULT 15
AAU74796
ID   AAU74796 standard; Protein; 183 AA.
XX
AC   AAU74796;
XX
DT   09-APR-2002  (first entry)
XX
DE   Mouse thrombospondin-2 (TSP-2) derived recombinant protein, 3TSR.
XX
KW   Thrombospondin-2; TSP-2; cytostatic; angiogenesis; vasotropic;
KW   vulnerary; neovascularisation; cell proliferation inhibitor; cancer;
KW   solid tumour; haemangioma; acoustic neuromas; neurofibroma; trachoma;
KW   pyogenic granulomas; rheumatoid arthritis; ocular angiogenic disease;
KW   retinopathy; psoriasis; macular degeneration; corneal graft rejection;
KW   neovascular glaucoma; retrolental fibroplasia; rebeosis; angiofibroma;
KW   Osler-Webber syndrome; myocardial angiogenesis; haemophiliac joints;
KW   plaque neovascularisation; telangiectasia; wound granulation;
KW   apoptosis; mutant; mutein; 3TSR.
XX
OS   Mus sp.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..5
FT                   /label= Vector_derived_peptide
FT   Protein         6..176
FT                   /note= "Residues 381-550 of mouse thrombospondin-2
FT                         (TSP-2)"
FT   Peptide         177..183
FT                   /label= Vector_derived_peptide
XX
PN   WO200191781-A2.
XX
PD   06-DEC-2001.
XX
PF   25-MAY-2001; 2001WO-US17250.
XX
PR   26-MAY-2000; 2000US-207994P.
XX
PA   (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
PI   Lawler JW;
XX
DR   WPI; 2002-106273/14.
XX
PT   Composition useful for treatment of cancer comprises cDNA encoding
PT   amino acids of human thrombospondin-1 or its conservative variant and a
PT   carrier  -
XX
PS   Example 2; Page -; 54pp; English.
XX
CC   The invention describes a composition comprising cDNA encoding fragments
CC   of human thrombospondin-1 (TSP -1), a type 1 repeat polypeptide and
CC   potent inhibitor of tumour growth and angiogenesis. The composition is
CC   useful for killing cancerous cells (preferably tumour); for reducing
CC   volume or inhibiting growth of a tumour (inhibiting neovascularisation in
CC   the tumour); for decreasing proliferation of tumour cells; in the
CC   treatment of diseases and conditions associated with angiogenic activity
CC   or misregulated growth and angiogenesis-mediated diseases such as cancer,
CC   solid tumour, tumour metastasis, benign tumour, (e.g. haemangioma,
CC   acoustic neuromas, neurofibromas, trachomas, and pyogenic granulomas),
CC   rheumatoid arthritis, psoriasis, ocular angiogenic diseases (e.g.
CC   diabetic retinopathy, retinopathy of prematurity, macular degeneration,
CC   corneal graft rejection, neovascular glaucoma, retrolental fibroplasias,
CC   rebeosis), Osler-Webber syndrome, myocardial angiogenesis,
CC   telangiectasia, plaque neovascularisation, haemophiliac joints,
CC   angiofibroma or wound granulation. The composition induces apoptosis and
CC   inhibits neovascularisation in the tumour cells. This amino acid sequence
CC   represents a recombinant protein, 3TSR, derived from mouse
CC   thrombospondin-2 (TSP-2) and containing the 3 type 1 repeat (TSR) regions
CC   of TSP-2, described in the method of the invention.
CC   Note: This sequence does not appear in the specification but has been
CC   created using the mouse thrombospondin-2 (TSP-2) wild type sequence
CC   (see AAU74786) from information given in the invention.
XX
SQ   Sequence   183 AA;

  Query Match             23.1%;  Score 57.5;  DB 23;  Length 183;
  Best Local Similarity   27.9%;  Pred. No. 10;
  Matches   17;  Conservative    6;  Mismatches   17;  Indels   21;  Gaps    3;

Qy       2 WYSWKRSSKSTGLGDTATR---------------EGCG----PSQSDGCPYNGRLTTVK 41
           |  |   | : |:|:  ||               :| |    | | | || :|| :
Db      68 WSPWSSCSVTCGVGN-VTRIRLCNSPVPQMGGKNCKGSGRETKPCQRDPCPIDGRWSPWS 126

Qy      42 P 42
           |
Db     127 P 127


Search completed: January 15, 2003, 08:17:06
Job time : 7.31507 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:16:02 ; Search time 9.09132 Seconds
(without alignments)
96.208 Million cell updates/sec

Title: US-09-079-819-50
Perfect score: 249
Sequence: 1 EWYSWKRSSKSTGLGDTATR.....PSQSDGCPYNGRLTTVKPRK 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	57	22.9	469	10	US-09-925-301-1279	Sequence 1279, Ap
2	55.5	22.3	939	10	US-09-854-845-16	Sequence 16, Appl
3	55.5	22.3	954	10	US-09-854-845-14	Sequence 14, Appl
4	55.5	22.3	1034	10	US-09-854-845-6	Sequence 6, Appli
5	55.5	22.3	1049	10	US-09-854-845-2	Sequence 2, Appli
6	55.5	22.3	1078	10	US-09-854-845-8	Sequence 8, Appli
7	55.5	22.3	1093	10	US-09-854-845-4	Sequence 4, Appli
8	55.5	22.3	1136	10	US-09-854-845-12	Sequence 12, Appl
9	55.5	22.3	1151	10	US-09-854-845-10	Sequence 10, Appl
10	55	22.1	5405	9	US-10-025-380-1116	Sequence 1116, Ap
11	55	22.1	5405	10	US-09-922-217-1116	Sequence 1116, Ap
12	54.5	21.9	118	9	US-10-001-857-187	Sequence 187, App
13	54.5	21.9	1588	9	US-10-000-512-2	Sequence 2, Appli
14	54	21.7	282	10	US-09-764-864-1174	Sequence 1174, Ap
15	53.5	21.5	800	10	US-09-764-853-800	Sequence 800, App
16	53.5	21.5	800	10	US-09-764-898-272	Sequence 272, App
17	53.5	21.5	877	10	US-09-764-898-200	Sequence 200, App
18	53	21.3	1156	10	US-09-757-716-1	Sequence 1, Appli
19	52	20.9	338	9	US-09-978-295A-119	Sequence 119, App
20	52	20.9	338	9	US-09-978-697-119	Sequence 119, App
21	52	20.9	338	9	US-09-978-192A-119	Sequence 119, App
22	52	20.9	338	9	US-09-999-832A-119	Sequence 119, App
23	52	20.9	338	9	US-09-978-189-119	Sequence 119, App
24	52	20.9	553	10	US-09-981-649A-6	Sequence 6, Appli
25	52	20.9	553	10	US-09-981-649A-24	Sequence 24, Appl
26	52	20.9	554	10	US-09-981-649A-30	Sequence 30, Appl
27	52	20.9	554	10	US-09-981-649A-32	Sequence 32, Appl
28	52	20.9	559	10	US-09-981-649A-28	Sequence 28, Appl
29	51.5	20.7	52	9	US-09-796-692-1529	Sequence 1529, Ap
30	51.5	20.7	178	10	US-09-764-853-469	Sequence 469, App
31	50.5	20.3	48	10	US-09-864-761-42865	Sequence 42865, A
32	50	20.1	81	10	US-09-764-860-399	Sequence 399, App
33	49.5	19.9	147	9	US-10-016-634A-115	Sequence 115, App
34	49	19.7	459	9	US-10-195-517-2	Sequence 2, Appli
35	49	19.7	459	10	US-09-741-152-2	Sequence 2, Appli
36	49	19.7	849	10	US-09-752-639-152	Sequence 152, App
37	49	19.7	849	10	US-09-984-198-152	Sequence 152, App
38	49	19.7	2491	9	US-10-000-789-2	Sequence 2, Appli
39	48.5	19.5	211	9	US-09-731-449-29	Sequence 29, Appl
40	48.5	19.5	258	9	US-09-731-449-28	Sequence 28, Appl
41	48.5	19.5	267	10	US-09-853-386-106	Sequence 106, App
42	48.5	19.5	277	10	US-09-864-761-46564	Sequence 46564, A
43	48.5	19.5	278	9	US-09-731-449-25	Sequence 25, Appl
44	48.5	19.5	499	10	US-09-735-169A-2	Sequence 2, Appli
45	48.5	19.5	499	10	US-09-735-171A-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-925-301-1279
; Sequence 1279, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1279
; LENGTH: 469
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (15)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (81)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1279

Query Match 22.9%; Score 57; DB 10; Length 469;
Best Local Similarity 41.4%; Pred. No. 11;
Matches 12; Conservative 2; Mismatches 7; Indels 8; Gaps 2;

```
Qy       4 SWKRSSKSTGLGDTATREGCGPSQSDGCP 32
           :|:    | | |     |||||    |||
Db      78 AWRXPGSSKGCG-----EGCGP---QGCP 98
```

RESULT 2
US-09-854-845-16
; Sequence 16, Application US/09854845
; Patent No. US20020098491A1

```
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleotid
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
;   LENGTH: 939
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-16

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 939;
  Best Local Similarity   27.8%;  Pred. No. 34;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |             :|||   | |   |
Db     531 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 584


RESULT 3
US-09-854-845-14
; Sequence 14, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleotid
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
;   LENGTH: 954
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-14

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 954;
  Best Local Similarity   27.8%;  Pred. No. 34;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |             :|||   | |   |
Db     531 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 584


RESULT 4
US-09-854-845-6
; Sequence 6, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleo
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 1034
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-6

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 1034;
  Best Local Similarity   27.8%;  Pred. No. 37;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |             :|||   | |   |
Db     626 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 679


RESULT 5
US-09-854-845-2
; Sequence 2, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleo
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1049
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-2

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 1049;
  Best Local Similarity   27.8%;  Pred. No. 38;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |             :|||   | |   |
Db     626 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 679


RESULT 6
US-09-854-845-8
; Sequence 8, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleo
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
```

```
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 1078
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-8

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 1078;
  Best Local Similarity   27.8%;  Pred. No. 39;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |              :|||   | |   |
Db     670 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 723


RESULT 7
US-09-854-845-4
; Sequence 4, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleotid
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1093
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-4

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 1093;
  Best Local Similarity   27.8%;  Pred. No. 40;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |              :|||   | |   |
Db     670 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 723


RESULT 8
US-09-854-845-12
; Sequence 12, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleotid
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12
;   LENGTH: 1136
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-12

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 1136;
  Best Local Similarity   27.8%;  Pred. No. 41;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |              :|||   | |   |
Db     728 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 781


RESULT 9
US-09-854-845-10
; Sequence 10, Application US/09854845
; Patent No. US20020098491A1
; GENERAL INFORMATION:
;  APPLICANT: Walke, D. Wade
;  APPLICANT:  Wang, Xiaoming
;  APPLICANT:  Scoville, John
;  APPLICANT:  Turner, C. Alexander Jr.
;  TITLE OF INVENTION: No. US20020098491A1el Human Semaphorin Homologs and Polynucleo
;  FILE REFERENCE: LEX-0177-USA
;  CURRENT APPLICATION NUMBER: US/09/854,845
;  CURRENT FILING DATE:  2001-05-14
;  PRIOR APPLICATION NUMBER: US 60/205,274
;  PRIOR FILING DATE: 2000-05-18
;  PRIOR APPLICATION NUMBER: US 60/208,893
;  PRIOR FILING DATE: 2000-06-02
;  NUMBER OF SEQ ID NOS: 50
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 1151
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-09-854-845-10

  Query Match             22.3%;  Score 55.5;  DB 10;  Length 1151;
  Best Local Similarity   27.8%;  Pred. No. 42;
  Matches   15;  Conservative    4;  Mismatches   22;  Indels   13;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQS------------DGCPYNGRLTTVKP 42
           | || : | : | |  : |  |    |              :|||   | |   |
Db     728 WGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTP 781


RESULT 10
US-10-025-380-1116
; Sequence 1116, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole L.
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick Thomas S.
;  APPLICANT:  Carter, Darrick
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C14
```

```
;  CURRENT APPLICATION NUMBER: US/10/025,380
;  CURRENT FILING DATE:  2001-12-19
;  NUMBER OF SEQ ID NOS: 1129
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1116
;   LENGTH: 5405
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-025-380-1116

  Query Match             22.1%;  Score 55;  DB 9;  Length 5405;
  Best Local Similarity   41.4%;  Pred. No. 2.5e+02;
  Matches   12;  Conservative    2;  Mismatches    7;  Indels    8;  Gaps    2;

Qy       4 SWKRSSKSTGLGDTATREGCGPSQSDGCP 32
           :|:    | | |    |||||   |||
Db    5014 AWRVPGSSKGCG-----EGCGP---QGCP 5034


RESULT 11
US-09-922-217-1116
; Sequence 1116, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1116
;   LENGTH: 5405
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-217-1116

  Query Match             22.1%;  Score 55;  DB 10;  Length 5405;
  Best Local Similarity   41.4%;  Pred. No. 2.5e+02;
  Matches   12;  Conservative    2;  Mismatches    7;  Indels    8;  Gaps    2;

Qy       4 SWKRSSKSTGLGDTATREGCGPSQSDGCP 32
           :|:    | | |    |||||   |||
Db    5014 AWRVPGSSKGCG-----EGCGP---QGCP 5034


RESULT 12
US-10-001-857-187
; Sequence 187, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
;  APPLICANT: Macina, Roberto
;  APPLICANT:  Recipon, Herve
;  APPLICANT:  Chen, Sei-Yu
;  APPLICANT:  Sun, Yongming
;  APPLICANT:  Liu, Chenghua
;  TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and Prot
;  FILE REFERENCE: DEX-0273
;  CURRENT APPLICATION NUMBER: US/10/001,857
;  CURRENT FILING DATE:  2001-11-20
;  PRIOR APPLICATION NUMBER: 60/252,054
;  PRIOR FILING DATE: 2000-11-20
;  NUMBER OF SEQ ID NOS: 208
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
;   LENGTH: 118
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-10-001-857-187

  Query Match             21.9%;  Score 54.5;  DB 9;  Length 118;
  Best Local Similarity   45.8%;  Pred. No. 5;
  Matches   11;  Conservative    3;  Mismatches    9;  Indels    1;  Gaps    1;

Qy      22 GCGPSQSDG-CPYNGRLTTVKPRK 44
           ||| ::  | ||: |    ||| |
Db      84 GCGGTEKWGVCPWGGAALLVKPEK 107


RESULT 13
US-10-000-512-2
; Sequence 2, Application US/10000512
; Patent No. US20020164699A1
; GENERAL INFORMATION:
;  APPLICANT: Shimkets, Richard A
;  APPLICANT:  Fernandes, Elma
;  TITLE OF INVENTION: POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME
;  FILE REFERENCE: 15966-556
;  CURRENT APPLICATION NUMBER: US/10/000,512
;  CURRENT FILING DATE:  2001-10-23
;  PRIOR APPLICATION NUMBER: 09/619,252
;  PRIOR FILING DATE: 2000-07-19
;  PRIOR APPLICATION NUMBER: 60/167,785
;  PRIOR FILING DATE: 1999-11-29
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1588
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-000-512-2

  Query Match             21.9%;  Score 54.5;  DB 9;  Length 1588;
  Best Local Similarity   37.1%;  Pred. No. 78;
  Matches   13;  Conservative    2;  Mismatches   15;  Indels    5;  Gaps    1;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           || :| | ||| | |     :          ||||
Db     971 EWSNWSRCSKSCGSGVKVRSKWLREK-----PYNG 1000


RESULT 14
US-09-764-864-1174
; Sequence 1174, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PTZ23
;  CURRENT APPLICATION NUMBER: US/09/764,864
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1792
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1174
;   LENGTH: 282
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (41)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (65)
```

```
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1174

  Query Match             21.7%;  Score 54;  DB 10;  Length 282;
  Best Local Similarity   31.4%;  Pred. No. 15;
  Matches   16;  Conservative    6;  Mismatches   21;  Indels    8;  Gaps    2;

Qy       2 WYSWKRSSKSTGLGD-----TATREGCGPSQSD---GCPYNGRLTTVKPRK 44
           :||||::   | | |      | |   | |  |     | :|  :   |:|
Db     198 YYSWKKTRSRTSLMDRQARKLANRHNQGDSDDDVEETHPMDGNDSDYDPKK 248


RESULT 15
US-09-764-853-800
; Sequence 800, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PJZ06
;  CURRENT APPLICATION NUMBER: US/09/764,853
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 939
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 800
;   LENGTH: 800
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (197)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (201)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (687)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-853-800

  Query Match             21.5%;  Score 53.5;  DB 10;  Length 800;
  Best Local Similarity   41.7%;  Pred. No. 50;
  Matches   15;  Conservative    0;  Mismatches   16;  Indels    5;  Gaps    2;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCP-YNGR 36
           | ||   | | | |      ||    ||  | | ||
Db      74 WASWSACSVSCGGGARQRTRGC----SDPVPQYGGR 105


Search completed: January 15, 2003, 08:23:04
Job time : 10.0913 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:27 ; Search time 3.26484 Seconds
(without alignments)
1295.596 Million cell updates/sec

Title: US-09-079-819-50
Perfect score: 249
Sequence: 1 EWYSWKRSSKSTGLGDTATR.....PSQSDGCPYNGRLTTVKPRK 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	68	27.3	417	2	A49129	neuroblast cell li
2	68	27.3	799	2	JH0797	castor protein - f
3	65	26.1	984	2	T00326	hypothetical prote
4	65	26.1	1522	2	T00028	brain-specific ang
5	64.5	25.9	1572	2	T00027	brain-specific ang
6	59	23.7	394	2	S36512	E2 protein - human
7	59	23.7	1360	2	T33922	hypothetical prote
8	58.5	23.5	1003	2	T13951	DNA topoisomerase
9	58.5	23.5	2533	2	T28675	alpha-51D immobili
10	58.5	23.5	2533	2	T28674	alpha-51D-immobili
11	57.5	23.1	509	2	AD0649	probable secreted
12	57.5	23.1	1172	2	A42587	thrombospondin 2 p
13	56	22.5	110	2	D72701	hypothetical prote
14	56	22.5	301	2	H70644	hypothetical prote
15	56	22.5	1444	2	T18856	angiogenesis inhib
16	55.5	22.3	488	2	S67744	hypothetical prote
17	54.5	21.9	189	2	B29795	circumsporozoite p
18	54.5	21.9	412	1	OZZQAF	circumsporozoite p
19	54.5	21.9	424	2	A54533	circumsporozoite p
20	54.5	21.9	1584	2	T00026	brain-specific ang
21	54	21.7	225	2	H69677	involved in polyke
22	54	21.7	438	1	S70602	cellulose 1,4-beta
23	54	21.7	1234	2	S52099	phospholipase C be
24	53.5	21.5	263	2	S13614	major allergen Lol
25	53.5	21.5	263	2	B37881	pollen allergen Lo
26	53.5	21.5	331	2	S45584	transcription acti
27	53.5	21.5	530	2	S33673	probable protein k
28	53.5	21.5	566	1	HMIVAT	hemagglutinin prec
29	53.5	21.5	1376	2	S63986	collagen alpha 5 c
30	53.5	21.5	1433	2	A46053	bullous pemphigoid
31	53	21.3	184	2	T46433	hypothetical prote
32	53	21.3	215	2	S34163	homeotic protein H
33	53	21.3	226	2	T43814	conserved hypothet
34	53	21.3	558	2	T17324	hypothetical prote
35	53	21.3	569	2	T19483	hypothetical prote
36	53	21.3	1327	2	T09402	immunoglobulin-lik
37	52	20.9	177	2	T19610	hypothetical prote
38	52	20.9	435	2	JC4590	salicylate 1-monoo
39	52	20.9	1221	2	A44978	DNA topoisomerase
40	51.5	20.7	121	1	A37331	secretory protein
41	51.5	20.7	388	2	A39756	circumsporozoite p
42	51.5	20.7	434	2	A39181	salicylate 1-monoo
43	51.5	20.7	788	2	T25061	hypothetical prote
44	51.5	20.7	951	2	T00260	hypothetical prote
45	51	20.5	292	2	T14627	hypothetical prote

ALIGNMENTS

RESULT 1
A49129
neuroblast cell lineage zinc finger protein ming - fruit fly (Drosophila melanogaster
C;Species: Drosophila melanogaster
C;Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 16-Feb-1997
C;Accession: A49129
R;Cui, X.; Doe, C.Q.
Development 116, 943-952, 1992
A;Title: ming is expressed in neuroblast sublineages and regulates gene expression in
A;Reference number: A49129; MUID:93202016; PMID:1339340
A;Accession: A49129
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-417 <CUI>
A;Experimental source: embryo
A;Note: sequence extracted from NCBI backbone (NCBIN:127816, NCBIP:127817)
C;Genetics:
A;Gene: FlyBase:cas
A;Cross-references: FlyBase:FBgn0004878

Query Match 27.3%; Score 68; DB 2; Length 417;
Best Local Similarity 26.3%; Pred. No. 1.1;
Matches 10; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

```
Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLT 38
           :|:  :: |   |    :: : | |:  :|| || : |
Db     155 KWHKKRKESLKLGFARFSSSDDCAPAYGEGCAYNWKQT 192
```

RESULT 2
JH0797
castor protein - fruit fly (Drosophila sp.)
C;Species: Drosophila sp.
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 31-Dec-2000
C;Accession: JH0797
R;Mellerick, D.M.; Kassis, J.A.; Zhang, S.D.; Odenwald, W.F.
Neuron 9, 789-803, 1992
A;Title: Castor encodes a novel zinc finger protein required for the development of a
A;Reference number: JH0797; MUID:93040222; PMID:1418995
A;Accession: JH0797
A;Molecule type: DNA
A;Residues: 1-799 <MEL>
A;Cross-references: GB:L04487
C;Comment: This protein is required for the development of embryonic CNS neurons.
C;Genetics:
A;Gene: FlyBase:cas
A;Cross-references: FlyBase:FBgn0004878
A;Introns: 284/3; 515/3; 660/3
C;Keywords: metal binding; zinc finger
F;236-242/Region: acidic
F;363-421/Region: zinc finger
F;422-482/Region: zinc finger

```
F;483-541/Region: zinc finger
F;542-600/Region: zinc finger

  Query Match             27.3%;  Score 68;  DB 2;  Length 799;
  Best Local Similarity   26.3%;  Pred. No. 2;
  Matches   10;  Conservative   10;  Mismatches   18;  Indels    0;  Gaps    0;

Qy      1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLT 38
          :|:  :: |   |    :: : | |:  :|| || : |
Db    400 KWHKKRKESLKLGFARFSSSDDCAPAYGEGCAYNWKQT 437


RESULT 3
T00326
hypothetical protein KIAA0550 - human
C;Species: Homo sapiens (man)
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C;Accession: T00326
R;Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A;Title: Prediction of the coding sequences of unidentified human genes. IX. The complet
A;Reference number: Z14086; MUID:98290545; PMID:9628581
A;Accession: T00326
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-984 <NAG>
A;Cross-references: EMBL:AB011122; NID:g3043623; PIDN:BAA25476.1; PID:g3043624
A;Experimental source: brain
C;Genetics:
A;Note: KIAA0550
C;Superfamily: thrombospondin type 1 repeat homology
F;344-398/Domain: thrombospondin type 1 repeat homology <THR3>

  Query Match             26.1%;  Score 65;  DB 2;  Length 984;
  Best Local Similarity   35.7%;  Pred. No. 5.6;
  Matches   15;  Conservative    1;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKP 42
          ||  |   | : | |       | | |  | |  |  |  ||
Db    350 EWSPWSLCSFTCGRGQRTRTRSCTPPQYGGRPCEGPETHHKP 391


RESULT 4
T00028
brain-specific angiogenesis inhibitor 3 - human
N;Alternate names: BAI3 protein
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C;Accession: T00028
R;Shiratsuchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.
Cytogenet. Cell Genet. 79, 103-108, 1997
A;Title: Cloning and characterization of BAI2 and BAI3, novel genes homologous to brain-
A;Reference number: Z14066; MUID:98194217; PMID:9533023
A;Accession: T00028
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1522 <SHI>
A;Cross-references: EMBL:AB005299; NID:g3021700; PIDN:BAA25363.1; PID:g3021701
A;Experimental source: brain
C;Genetics:
A;Gene: GDB:BAI3
A;Cross-references: GDB:9838090; OMIM:602684
A;Map position: 6q12-6q12
C;Superfamily: thrombospondin type 1 repeat homology
F;344-398/Domain: thrombospondin type 1 repeat homology <THR3>

  Query Match             26.1%;  Score 65;  DB 2;  Length 1522;
  Best Local Similarity   35.7%;  Pred. No. 8.3;
  Matches   15;  Conservative    1;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKP 42
          ||  |   | : | |       | | |  | |  |  |  ||
Db    350 EWSPWSLCSFTCGRGQRTRTRSCTPPQYGGRPCEGPETHHKP 391


RESULT 5
T00027
brain-specific angiogenesis inhibitor 2 - human
N;Alternate names: BAI2 protein
C;Species: Homo sapiens (man)
C;Date: 22-Jan-1999 #sequence_revision 22-Jan-1999 #text_change 21-Jul-2000
C;Accession: T00027
R;Shiratsuchi, T.; Nishimori, H.; Ichise, H.; Nakamura, Y.; Tokino, T.
Cytogenet. Cell Genet. 79, 103-108, 1997
A;Title: Cloning and characterization of BAI2 and BAI3, novel genes homologous to bra
A;Reference number: Z14066; MUID:98194217; PMID:9533023
A;Accession: T00027
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1572 <SHI>
A;Cross-references: EMBL:AB005298; NID:g3021698; PIDN:BAA25362.1; PID:g3021699
A;Experimental source: brain
C;Genetics:
A;Gene: GDB:BAI2
A;Cross-references: GDB:9838089; OMIM:602683
A;Map position: 1p35-1p35

  Query Match             25.9%;  Score 64.5;  DB 2;  Length 1572;
  Best Local Similarity   28.3%;  Pred. No. 9.9;
  Matches   15;  Conservative    3;  Mismatches   18;  Indels   17;  Gaps    1;

Qy      1 EWYSWKRSSKSTGLGDTATREGCGPSQSDG----------------CPYNGR 36
          || ||   |:| | |  :    | | |  |                 ||  |:
Db    357 EWGSWSLCSRSCGRGSRSRMRTCVPPQHGGKACEGPELQTKLCSMAACPVEGQ 409


RESULT 6
S36512
E2 protein - human papillomavirus type 32
C;Species: human papillomavirus type 32
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
C;Accession: S36512
R;Delius, H.; Hofmann, B.
submitted to the EMBL Data Library, August 1993
A;Description: Primer-directed sequencing of human papillomavirus types.
A;Reference number: S36469
A;Accession: S36512
A;Molecule type: DNA
A;Residues: 1-394 <DEL>
A;Cross-references: EMBL:X74475; NID:g396981; PIDN:CAA52552.1; PID:g396985
C;Superfamily: papillomavirus E2 protein
C;Keywords: DNA binding; early protein; transcription regulation

  Query Match             23.7%;  Score 59;  DB 2;  Length 394;
  Best Local Similarity   42.9%;  Pred. No. 13;
  Matches   12;  Conservative    1;  Mismatches   15;  Indels    0;  Gaps    0;

Qy     17 TATREGCGPSQSDGCPYNGRLTTVKPRK 44
          | ||  | |  :||  |     |  |||
Db    233 TITRSNCDPDGTDGILYKDPTPTTPPRK 260


RESULT 7
T33922
hypothetical protein Y8A9A.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33922
R;Courtney, L.; Langston, Y.; Drone, K.; Mead, K.
submitted to the EMBL Data Library, February 1999
A;Description: The sequence of C. elegans cosmid Y8A9A.
A;Reference number: Z21439
A;Accession: T33922
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1360 <COU>
```

```
A;Cross-references: EMBL:AF125461; PIDN:AAD12852.1; GSPDB:GN00020; CESP:Y8A9A.2
A;Experimental source: strain Bristol N2; clone Y8A9A
C;Genetics:
A;Gene: CESP:Y8A9A.2
A;Map position: 2
A;Introns: 43/3; 69/3; 134/3; 174/1; 604/2; 643/3; 718/3; 787/3; 848/2; 944/2; 1044/2; 1

  Query Match             23.7%;  Score 59;  DB 2;  Length 1360;
  Best Local Similarity   36.8%;  Pred. No. 41;
  Matches   14;  Conservative    3;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLT 38
           :| :|   | : |   | ||     |   |||  | ||
Db    1207 DWGAWTTCSATCGGCGTQTRSRTCASAPYGCPCTGDLT 1244


RESULT 8
T13951
DNA topoisomerase (EC 5.99.1.2) III - mouse
C;Species: Mus musculus (house mouse)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T13951
R;Seki, T.; Seki, M.; Katada, T.; Enomoto, T.
Biochim. Biophys. Acta 1396, 127-131, 1998
A;Title: Isolation of a cDNA encoding mouse DNA topoisomerase III which is highly expres
A;Reference number: Z17829; MUID:98201702; PMID:9540825
A;Accession: T13951
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1003 <SEK>
A;Cross-references: EMBL:AB006074; NID:g3061307; PIDN:BAA25662.1; PID:g3061308
C;Genetics:
A;Gene: mTOP3
C;Keywords: isomerase

  Query Match             23.5%;  Score 58.5;  DB 2;  Length 1003;
  Best Local Similarity   45.2%;  Pred. No. 36;
  Matches   14;  Conservative    2;  Mismatches   14;  Indels    1;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCP 32
           :: |  || ||| | | |     |  | |||
Db     847 FFLWADSSHSTG-GGTPTSASGPPGSSVGCP 876


RESULT 9
T28675
alpha-51D immobilization antigen - Paramecium tetraurelia
C;Species: Paramecium tetraurelia
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C;Accession: T28675
R;Schwegmann, K.J.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z20506
A;Accession: T28675
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2533 <SCH>
A;Cross-references: EMBL:X96400; PIDN:CAA65264.1
C;Genetics:
A;Gene: alpha-51D
A;Genetic code: SGC5
A;Introns: 280/3; 538/2; 1248/2
C;Superfamily: G surface protein

  Query Match             23.5%;  Score 58.5;  DB 2;  Length 2533;
  Best Local Similarity   41.2%;  Pred. No. 82;
  Matches   14;  Conservative    3;  Mismatches   16;  Indels    1;  Gaps    1;

Qy      10 KSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPR 43
           |: | :  |   || |  ||| |||    | |:
Db     700 KTCKLAENQTDGSCG-SFLDGCVYNGNAGCVDPK 732


RESULT 10
T28674
alpha-51D-immobilization antigen - Paramecium tetraurelia
C;Species: Paramecium tetraurelia
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C;Accession: T28674
R;Schmidt, H.J.
submitted to the EMBL Data Library, March 1995
A;Reference number: Z20505
A;Accession: T28674
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2533 <SCH>
A;Cross-references: EMBL:X85135; NID:g728634; PID:g728635; PIDN:CAA59447.1
C;Genetics:
A;Genetic code: SGC5
A;Note: alpha-51D
C;Superfamily: G surface protein

  Query Match             23.5%;  Score 58.5;  DB 2;  Length 2533;
  Best Local Similarity   41.2%;  Pred. No. 82;
  Matches   14;  Conservative    3;  Mismatches   16;  Indels    1;  Gaps    1;

Qy      10 KSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPR 43
           |: | :  |   || |  ||| |||    | |:
Db     700 KTCKLAENQTDGSCG-SFLDGCVYNGNAGCVDPK 732


RESULT 11
AD0649
probable secreted protein STY1292 [imported] - Salmonella enterica subsp. enterica se
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C;Accession: AD0649
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A;Reference number: AB0502; PMID:11677608
A;Accession: AD0649
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-509 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08375.1; PID:g16502420; GSPDB:GN00176
C;Genetics:
A;Gene: STY1292

  Query Match             23.1%;  Score 57.5;  DB 2;  Length 509;
  Best Local Similarity   37.0%;  Pred. No. 25;
  Matches   17;  Conservative    6;  Mismatches   16;  Indels    7;  Gaps    2;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPS--QSDGCPYNGRLTTVKPRK 44
           |||       |:  |||::|:|  | :    || | | ::|     ||
Db     112 EWY-----RKAAELGDSSTQEILGDAYMYGDGFPQNTQLALEWYRK 152


RESULT 12
A42587
thrombospondin 2 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999
C;Accession: A42587; A39851
R;Laherty, C.D.; O'Rourke, K.; Wolf, F.W.; Katz, R.; Seldin, M.F.; Dixit, V.M.
J. Biol. Chem. 267, 3274-3281, 1992
A;Title: Characterization of mouse thrombospondin 2 sequence and expression during ce
A;Reference number: A42587; MUID:92147683; PMID:1371115
A;Accession: A42587
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1172 <LAH>
```

```
A;Cross-references: GB:L07803; GB:M87275; NID:g340421; PIDN:AAA53064.1; PID:g567241
A;Note: sequence extracted from NCBI backbone (NCBIP:81502)
R;Bornstein, P.; O'Rourke, K.; Wikstrom, K.; Wolf, F.W.; Katz, R.; Li, P.; Dixit, V.M.
J. Biol. Chem. 266, 12821-12824, 1991
A;Title: A second, expressed thrombospondin gene (Thbs2) exists in the mouse genome.
A;Reference number: A39851; MUID:91302287; PMID:1712771
A;Accession: A39851
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-873 <BOR>
A;Cross-references: GB:M64866; NID:g201994; PIDN:AAA40432.1; PID:g201995
C;Superfamily: thrombospondin 1; EGF homology; thrombospondin type 1 repeat homology; vc
C;Keywords: calcium binding; glycoprotein
F;319-377/Domain: von Willebrand factor type C repeat homology <VWC>
F;380-431/Domain: thrombospondin type 1 repeat homology <THR1>
F;436-492/Domain: thrombospondin type 1 repeat homology <THR2>
F;493-549/Domain: thrombospondin type 1 repeat homology <THR3>
F;553-588/Domain: EGF homology <EGF1>
F;652-691/Domain: EGF homology <EGF>

  Query Match             23.1%;  Score 57.5;  DB 2;  Length 1172;
  Best Local Similarity   27.9%;  Pred. No. 54;
  Matches   17;  Conservative    6;  Mismatches   17;  Indels   21;  Gaps    3;

Qy       2 WYSWKRSSKSTGLGDTATR----------------EGCG----PSQSDGCPYNGRLTTVK 41
           |  |    | : |:|:  ||                :| |    | | | || :|| :
Db     443 WSPWSSCSVTCGVGN-VTRIRLCNSPVPQMGGKNCKGSGRETKPCQRDPCPIDGRWSPWS 501

Qy      42 P 42
           |
Db     502 P 502


RESULT 13
D72701
hypothetical protein APE1027 - Aeropyrum pernix (strain K1)
C;Species: Aeropyrum pernix
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C;Accession: D72701
R;Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A;Reference number: A72450; MUID:99310339; PMID:10382966
A;Accession: D72701
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-110 <KAW>
A;Cross-references: DDBJ:AP000060; NID:g5104188; PIDN:BAA80012.1; PID:d1043798; PID:g510
A;Experimental source: strain K1
C;Genetics:
A;Gene: APE1027
C;Superfamily: Aeropyrum pernix hypothetical protein APE1027

  Query Match             22.5%;  Score 56;  DB 2;  Length 110;
  Best Local Similarity   39.3%;  Pred. No. 9.7;
  Matches   11;  Conservative    4;  Mismatches   13;  Indels    0;  Gaps    0;

Qy       4 SWKRSSKSTGLGDTATREGCGPSQSDGC 31
           :|:||| | |:| :   | |  |    |
Db      53 TWRRSSASAGVGGSTWLEPCWASLGCSC 80


RESULT 14
H70644
hypothetical protein Rv0725c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70644
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70644
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-301 <COL>
A;Cross-references: GB:Z84395; GB:AL123456; NID:g3261698; PIDN:CAB06449.1; PID:g18061
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0725c
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv0731c

  Query Match             22.5%;  Score 56;  DB 2;  Length 301;
  Best Local Similarity   30.9%;  Pred. No. 24;
  Matches   17;  Conservative    8;  Mismatches   18;  Indels   12;  Gaps    4;

Qy       2 WYS-WKR----SSKSTGLGDTATREG-----CGPSQSDG--CPYNGRLTTVKPRK 44
           | | |:|    :: :|     | | |     |||: |    ||:    ||::  |
Db     240 WTSTWRRWCIPANAATSSTTCAPRAGTLRAQCGPTYSGAMVCPFPPHTTTIRSAK 294


RESULT 15
T18856
angiogenesis inhibitor homolog - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T18856; T24653
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19031
A;Accession: T18856
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1444 <WIL>
A;Cross-references: EMBL:Z50004; PIDN:CAA90293.1; GSPDB:GN00028; CESP:C02B4.1
A;Experimental source: clone C02B4
R;McMurray, A.
submitted to the EMBL Data Library, July 1995
A;Reference number: Z19917
A;Accession: T24653
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1444 <WI2>
A;Cross-references: EMBL:Z50006; PIDN:CAA90302.1; GSPDB:GN00028; CESP:C02B4.1
A;Experimental source: clone T07C5
C;Genetics:
A;Gene: CESP:C02B4.1
A;Map position: X
A;Introns: 25/3; 70/3; 96/3; 139/3; 187/1; 234/2; 282/3; 376/2; 422/2; 478/3; 509/3;

  Query Match             22.5%;  Score 56;  DB 2;  Length 1444;
  Best Local Similarity   52.4%;  Pred. No. 1e+02;
  Matches   11;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATRE 21
           || ||   | | |||    ||
Db     695 EWSSWSECSVSCGLGGREVRE 715


Search completed: January 15, 2003, 08:18:18
Job time : 6.26484 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:22 ; Search time 0.954338 Seconds
(without alignments)
1912.277 Million cell updates/sec

Title: US-09-079-819-50
Perfect score: 249
Sequence: 1 EWYSWKRSSKSTGLGDTATR.....PSQSDGCPYNGRLTTVKPRK 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	65	26.1	1522	1	BAI3_HUMAN	O60242 homo sapien
2	64.5	25.9	1572	1	BAI2_HUMAN	O60241 homo sapien
3	59	23.7	394	1	VE2_HPV32	P36791 human papil
4	58.5	23.5	1003	1	TP3A_MOUSE	O70157 mus musculu
5	57.5	23.1	1172	1	TSP2_MOUSE	Q03350 mus musculu
6	56.5	22.7	1093	1	SM5B_MOUSE	Q60519 mus musculu
7	55	22.1	867	1	SSPO_BOVIN	P98167 bos taurus
8	54.5	21.9	315	1	CSP_PLAFL	P05691 plasmodium
9	54.5	21.9	412	1	CSP_PLAFA	P02893 plasmodium
10	54.5	21.9	424	1	CSP_PLAFT	P13814 plasmodium
11	54.5	21.9	1584	1	BAI1_HUMAN	O14514 homo sapien
12	54	21.7	438	1	GUX3_AGABI	P49075 agaricus bi
13	53.5	21.5	263	1	MPL1_LOLPR	P14946 lolium pere
14	53.5	21.5	269	1	MPA1_PHAAQ	Q41260 phalaris aq
15	53.5	21.5	530	1	PILS_PSEAE	P33639 pseudomonas
16	53.5	21.5	565	1	HEMA_IATKO	P16060 influenza a
17	53.5	21.5	1173	1	TSP1_XENLA	P35448 xenopus lae
18	52	20.9	435	1	NHG2_PSEPU	Q53552 pseudomonas
19	52	20.9	1221	1	TOP2_TRYBB	P12531 trypanosoma
20	51.5	20.7	388	1	CSP_PLARE	P26694 plasmodium
21	51.5	20.7	433	1	NHG1_PSEPU	P23262 pseudomonas
22	51.5	20.7	439	1	XP2_XENLA	P17437 xenopus lae
23	51.5	20.7	564	1	HEMA_IAMAP	P87506 influenza a
24	51	20.5	296	1	NIFH_ALCFA	Q44044 alcaligenes
25	51	20.5	423	1	MKR2_SERQU	Q9dd48 seriola qui
26	51	20.5	437	1	PROP_MOUSE	P11680 mus musculu
27	50.5	20.3	1435	1	Y194_HUMAN	Q12766 homo sapien
28	50	20.1	397	1	CSP_PLAFO	P19597 plasmodium
29	50	20.1	519	1	DPYS_RAT	Q63150 rattus norv
30	50	20.1	551	1	YVD3_CAEEL	P55114 caenorhabdi
31	50	20.1	777	1	RTN1_RAT	Q64548 rattus norv
32	49.5	19.9	206	1	UMP3_ARATH	Q9lev3 arabidopsis
33	49.5	19.9	566	1	HEMA_IADIR	P11132 influenza a
34	49.5	19.9	566	1	HEMA_IATKI	P11135 influenza a
35	49.5	19.9	660	1	VE1_HPV29	P50764 human papil
36	49.5	19.9	799	1	ITBN_DROME	Q27591 drosophila
37	49.5	19.9	1628	1	NAGH_CLOPE	P26831 clostridium
38	49	19.7	129	1	VAL2_TMOV	Q06658 tomato mott
39	49	19.7	281	1	YH71_AERPE	Q9yb24 aeropyrum p
40	49	19.7	292	1	NIFH_KLEPN	P00458 klebsiella
41	49	19.7	303	1	BNK_DROME	P40794 drosophila
42	49	19.7	339	1	HXD9_MOUSE	P28357 mus musculu
43	49	19.7	359	1	GUN1_STRSQ	P13933 streptomyce
44	49	19.7	392	1	NADD_UREPA	Q9pq21 ureaplasma
45	49	19.7	461	1	VE2_HPV09	P36780 human papil

ALIGNMENTS

```
RESULT 1
BAI3_HUMAN
ID   BAI3_HUMAN     STANDARD;      PRT;  1522 AA.
AC   O60242; O60297;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Brain-specific angiogenesis inhibitor 3 precursor.
GN   BAI3 OR KIAA0550.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (LONG ISOFORM).
RC   TISSUE=Fetal brain;
RX   MEDLINE=98194217; PubMed=9533023;
RA   Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;
RT   "Cloning and characterization of BAI2 and BAI3, novel genes homologous
RT   to brain-specific angiogenesis inhibitor 1 (BAI1).";
RL   Cytogenet. Cell Genet. 79:103-108(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (SHORT ISOFORM).
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
CC   -!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION AND
CC       SUPPRESSION OF GLIOBLASTOMA.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC       SHORT FORM; MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN
CC       HEART. REDUCED EXPRESSION IS OBSERVED IN SOME GLIOBLASTOMA CELL
CC       LINES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 1 CUB DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB005299; BAA25363.1; -.
DR   EMBL; AB011122; BAA25476.1; -.
DR   Genew; HGNC:945; BAI3.
DR   MIM; 602684; -.
```

```
DR   InterPro; IPR000859; CUB_domain.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR001879; hormn_receptor.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00090; tsp_1; 4.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00209; TSP1; 4.
DR   PROSITE; PS01180; CUB; 1.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50092; TSP1; 4.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Repeat; Alternative splicing.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1522       BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 3.
FT   DOMAIN       25    880       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    881    901       1 (POTENTIAL).
FT   DOMAIN      902    910       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    911    931       2 (POTENTIAL).
FT   DOMAIN      932    939       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    940    960       3 (POTENTIAL).
FT   DOMAIN      961    981       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    982   1002       4 (POTENTIAL).
FT   DOMAIN     1003   1023       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1024   1044       5 (POTENTIAL).
FT   DOMAIN     1045   1098       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1099   1119       6 (POTENTIAL).
FT   DOMAIN     1120   1125       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1126   1146       7 (POTENTIAL).
FT   DOMAIN     1147   1522       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       30    159       CUB.
FT   DOMAIN      291    344       TSP TYPE-1 1.
FT   DOMAIN      345    399       TSP TYPE-1 2.
FT   DOMAIN      400    454       TSP TYPE-1 3.
FT   DOMAIN      455    509       TSP TYPE-1 4.
FT   DOMAIN      816    868       GPS.
FT   DOMAIN      942    945       POLY-THR.
FT   DOMAIN     1173   1176       POLY-SER.
FT   CARBOHYD     51     51       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     54     54       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     82     82       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    105    105       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    241    241       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    337    337       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    540    540       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    625    625       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    779    779       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    812    812       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    828    828       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    937    937       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    643    665       MISSING (IN SHORT ISOFORM).
FT   VARSPLIC    990   1007       LPALVVATSVGFTRTKGY -> KHIDIPFHALLKMTLNTH
FT                                (IN SHORT ISOFORM).
FT   VARSPLIC   1008   1522       MISSING (IN SHORT ISOFORM).
SQ   SEQUENCE   1522 AA;  171490 MW;  D22D0A5D4BB62502 CRC64;

  Query Match           26.1%;  Score 65;  DB 1;  Length 1522;
  Best Local Similarity  35.7%;  Pred. No. 1.8;
  Matches   15;  Conservative    1;  Mismatches   26;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKP 42
           ||  |   | : | |       | | |  | |  |  |  ||
Db     350 EWSPWSLCSFTCGRGQRTRTRSCTPPQYGGRPCEGPETHHKP 391
```

```
RESULT 2
BAI2_HUMAN
ID   BAI2_HUMAN     STANDARD;      PRT;  1572 AA.
AC   O60241;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Brain-specific angiogenesis inhibitor 2 precursor.
GN   BAI2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98194217; PubMed=9533023;
RA   Shiratsuchi T., Nishimori H., Ichise H., Nakamura Y., Tokino T.;
RT   "Cloning and characterization of BAI2 and BAI3, novel genes homologous
RT   to brain-specific angiogenesis inhibitor 1 (BAI1).";
RL   Cytogenet. Cell Genet. 79:103-108(1997).
CC   -!- FUNCTION: MIGHT BE INVOLVED IN ANGIOGENESIS INHIBITION.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN BRAIN. ALSO DETECTED IN
CC       HEART, THYMUS, SKELETAL MUSCLE, AND DIFFERENT CELL LINES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AB005298; BAA25362.1; -.
DR   Genew; HGNC:944; BAI2.
DR   MIM; 602683; -.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR001879; hormn_receptor.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00090; tsp_1; 4.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00209; TSP1; 4.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50092; TSP1; 4.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
KW   Repeat.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   1572       BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 2.
FT   DOMAIN       21    924       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    925    945       1 (POTENTIAL).
FT   DOMAIN      946    953       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    954    974       2 (POTENTIAL).
FT   DOMAIN      975    982       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    983   1003       3 (POTENTIAL).
FT   DOMAIN     1004   1024       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1025   1045       4 (POTENTIAL).
FT   DOMAIN     1046   1066       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1067   1087       5 (POTENTIAL).
FT   DOMAIN     1088   1141       CYTOPLASMIC (POTENTIAL).
```

```
FT   TRANSMEM   1142   1162       6 (POTENTIAL).
FT   DOMAIN     1163   1168       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1169   1189       7 (POTENTIAL).
FT   DOMAIN     1190   1572       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      297    351       TSP TYPE-1 1.
FT   DOMAIN      352    406       TSP TYPE-1 2.
FT   DOMAIN      407    461       TSP TYPE-1 3.
FT   DOMAIN      463    517       TSP TYPE-1 4.
FT   DOMAIN      859    911       GPS.
FT   DOMAIN      117    122       POLY-GLU.
FT   DOMAIN      177    180       POLY-ASN.
FT   DOMAIN      222    225       POLY-THR.
FT   DOMAIN     1303   1306       POLY-PRO.
FT   DOMAIN     1352   1358       POLY-GLY.
FT   DOMAIN     1413   1418       POLY-PRO.
FT   CARBOHYD     94     94       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    179    179       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    180    180       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    344    344       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    633    633       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    855    855       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1572 AA;  171140 MW;  A9775645B77BC285 CRC64;

  Query Match             25.9%;  Score 64.5;  DB 1;  Length 1572;
  Best Local Similarity   28.3%;  Pred. No. 2.2;
  Matches   15;  Conservative    3;  Mismatches   18;  Indels   17;  Gaps    1;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDG----------------CPYNGR 36
           || ||   |:| | |  :    | | |  |                 ||  |:
Db     357 EWGSWSLCSRSCGRGSRSRMRTCVPPQHGGKACEGPELQTKLCSMAACPVEGQ 409


RESULT 3
VE2_HPV32
ID   VE2_HPV32      STANDARD;      PRT;   394 AA.
AC   P36791;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Regulatory protein E2.
GN   E2.
OS   Human papillomavirus type 32.
OC   Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC   Papillomavirus.
OX   NCBI_TaxID=10612;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94265501; PubMed=8205838;
RA   Delius H., Hofmann B.;
RT   "Primer-directed sequencing of human papillomavirus types.";
RL   Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC   -!- FUNCTION: E2 REGULATES VIRAL TRANSCRIPTION AND DNA REPLICATION.
CC       IT BINDS TO THE E2RE RESPONSE ELEMENT (5'-ACCNNNNNNGGT-3') PRESENT
CC       IN MULTIPLE COPIES IN THE REGULATORY REGION. IT CAN EITHER
CC       ACTIVATE OR REPRESS TRANSCRIPTION DEPENDING OF E2RE'S POSITION
CC       WITH REGARDS TO PROXIMAL PROMOTER ELEMENTS. REPRESSION OCCURS
CC       BY STERICALLY HINDERING THE ASSEMBLY OF THE TRANSCRIPTION
CC       INITIATION COMPLEX. THE E1-E2 COMPLEX BINDS TO THE ORIGIN OF DNA
CC       REPLICATION.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X74475; CAA52552.1; -.
DR   PIR; S36512; S36512.
DR   HSSP; P17383; 1DHM.
DR   InterPro; IPR000427; E2_C.
DR   InterPro; IPR001866; E2_N.
DR   Pfam; PF00508; E2_N; 1.
DR   Pfam; PF00511; E2_C; 1.
DR   ProDom; PD000672; E2_C; 1.
DR   ProDom; PD000678; E2_N; 1.
KW   Early protein; Transcription regulation; Activator; DNA-binding;
KW   Trans-acting factor; DNA replication; Repressor; Nuclear protein.
SQ   SEQUENCE   394 AA;  45038 MW;  113C46119C2265E7 CRC64;

  Query Match             23.7%;  Score 59;  DB 1;  Length 394;
  Best Local Similarity   42.9%;  Pred. No. 2.7;
  Matches   12;  Conservative    1;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      17 TATREGCGPSQSDGCPYNGRLTTVKPRK 44
           | ||  | |  :||  |     |  |||
Db     233 TITRSNCDPDGTDGILYKDPTPTTPPRK 260


RESULT 4
TP3A_MOUSE
ID   TP3A_MOUSE     STANDARD;      PRT;  1003 AA.
AC   O70157;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   DNA topoisomerase III alpha (EC 5.99.1.2).
GN   TOP3A OR TOP3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Testis;
RX   MEDLINE=98201702; PubMed=9540825;
RA   Seki T., Seki M., Katada T., Enomoto T.;
RT   "Isolation of a cDNA encoding mouse DNA topoisomerase III which is
RT   highly expressed at the mRNA level in the testis.";
RL   Biochim. Biophys. Acta 1396:127-131(1998).
CC   -!- FUNCTION: REDUCES THE NUMBER OF SUPERCOILS IN A HIGHLY NEGATIVELY
CC       SUPERCOILED DNA.
CC   -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC       DNA, followed by passage and rejoining.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS.
CC   -!- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB006074; BAA25662.1; -.
DR   MGD; MGI:1197527; Top3a.
DR   InterPro; IPR002936; DNAprim_toprim.
DR   InterPro; IPR003601; DNAtopI_ATP_bind.
DR   InterPro; IPR003602; DNAtopI_DNA_bind.
DR   InterPro; IPR000380; Prok_tpisomrase.
DR   InterPro; IPR001878; Znf_CCHC.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   Pfam; PF01131; Topoisom_bac; 1.
DR   Pfam; PF01396; zf-C4_Topoisom; 1.
DR   Pfam; PF01751; Toprim; 1.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   PRINTS; PR00417; PRTPISMRASEI.
DR   SMART; SM00437; TOP1Ac; 1.
DR   SMART; SM00436; TOP1Bc; 1.
```

```
DR   SMART; SM00493; TOPRIM; 1.
DR   SMART; SM00343; ZnF_C2HC; 1.
DR   PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW   Isomerase; Topoisomerase; DNA-binding; Repeat; Zinc-finger.
FT   ACT_SITE    362    362       DNA CLEAVAGE (BY SIMILARITY).
FT   DOMAIN      658    685       C4-TYPE (POTENTIAL).
FT   DOMAIN      814    925       2 X 27 AA APPROXIMATE REPEATS.
FT   REPEAT      814    841       1.
FT   REPEAT      898    925       2.
SQ   SEQUENCE   1003 AA;  112358 MW;  254C738E746EE495 CRC64;

  Query Match            23.5%;  Score 58.5;  DB 1;  Length 1003;
  Best Local Similarity   45.2%;  Pred. No. 8.2;
  Matches   14;  Conservative    2;  Mismatches   14;  Indels    1;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCP 32
           :: |  || ||| | | |     |  | |||
Db     847 FFLWADSSHSTG-GGTPTSASGPPGSSVGCP 876


RESULT 5
TSP2_MOUSE
ID   TSP2_MOUSE     STANDARD;      PRT;  1172 AA.
AC   Q03350;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Thrombospondin 2 precursor.
GN   THBS2 OR TSP2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92147683; PubMed=1371115;
RA   Laherty C.D., O'Rourke K., Wolf F.W., Katz R., Seldin M.F.,
RA   Dixit V.M.;
RT   "Characterization of mouse thrombospondin 2 sequence and expression
RT   during cell growth and development.";
RL   J. Biol. Chem. 267:3274-3281(1992).
RN   [2]
RP   SEQUENCE OF 1-873 FROM N.A.
RX   MEDLINE=91302287; PubMed=1712771;
RA   Bornstein P., O'Rourke K., Wikstrom K., Wolf F.W., Katz R., Li P.,
RA   Dixit V.M.;
RT   "A second, expressed thrombospondin gene (Thbs2) exists in the mouse
RT   genome.";
RL   J. Biol. Chem. 266:12821-12824(1991).
CC   -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
CC       CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
CC       LAMININ AND TYPE V COLLAGEN.
CC   -!- SUBUNIT: HOMOTRIMER; DISULFIDE-LINKED.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L07803; AAA53064.1; -.
DR   EMBL; M64866; AAA40432.1; -.
DR   PIR; A42587; A42587.
DR   PIR; A39851; A39851.
DR   HSSP; P00740; 1EDM.
DR   MGD; MGI:98738; Thbs2.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWF_C.
DR   InterPro; IPR003367; tsp_3.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF02412; tsp_3; 9.
DR   SMART; SM00181; EGF; 3.
DR   SMART; SM00209; TSP1; 3.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS01208; VWFC; 1.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
KW   EGF-like domain; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19   1172       THROMBOSPONDIN 2.
FT   DOMAIN       19    232       HEPARIN-BINDING (POTENTIAL).
FT   DOMAIN      318    375       VWFC.
FT   DOMAIN      381    432       TSP TYPE-1 1.
FT   DOMAIN      437    493       TSP TYPE-1 2.
FT   DOMAIN      494    548       TSP TYPE-1 3.
FT   DOMAIN      549    589       EGF-LIKE 1.
FT   DOMAIN      590    647       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN      648    692       EGF-LIKE 3.
FT   DOMAIN      725    760       TSP TYPE-3 1.
FT   DOMAIN      761    783       TSP TYPE-3 2.
FT   DOMAIN      784    819       TSP TYPE-3 3.
FT   DOMAIN      820    842       TSP TYPE-3 4.
FT   DOMAIN      843    880       TSP TYPE-3 5.
FT   DOMAIN      881    916       TSP TYPE-3 6.
FT   DOMAIN      917    952       TSP TYPE-3 7.
FT   DOMAIN      953   1172       C-TERMINAL.
FT   SITE        928    930       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    266    266       INTERCHAIN (PROBABLE).
FT   DISULFID    270    270       INTERCHAIN (PROBABLE).
FT   DISULFID    553    564       BY SIMILARITY.
FT   DISULFID    558    574       BY SIMILARITY.
FT   DISULFID    577    588       BY SIMILARITY.
FT   DISULFID    594    610       BY SIMILARITY.
FT   DISULFID    601    619       BY SIMILARITY.
FT   DISULFID    622    646       BY SIMILARITY.
FT   DISULFID    652    665       BY SIMILARITY.
FT   DISULFID    659    678       BY SIMILARITY.
FT   DISULFID    680    691       BY SIMILARITY.
FT   CARBOHYD    151    151       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    316    316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    457    457       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    584    584       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    710    710       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1069   1069       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1172 AA;  129911 MW;  7CE8E4E8599822AB CRC64;

  Query Match            23.1%;  Score 57.5;  DB 1;  Length 1172;
  Best Local Similarity   27.9%;  Pred. No. 13;
  Matches   17;  Conservative    6;  Mismatches   17;  Indels   21;  Gaps    3;

Qy       2 WYSWKRSSKSTGLGDTATR---------------EGCG----PSQSDGCPYNGRLTTVK 41
           |  |   | : |:|:  ||               :| |    | | | || :|| :
Db     443 WSPWSSCSVTCGVGN-VTRIRLCNSPVPQMGGKNCKGSGRETKPCQRDPCPIDGRWSPWS 501

Qy      42 P 42
           |
Db     502 P 502
```

```
RESULT 6
SM5B_MOUSE
ID   SM5B_MOUSE     STANDARD;      PRT;  1093 AA.
AC   Q60519;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Semaphorin 5B precursor (Semaphorin G) (Sema G).
GN   SEMA5B OR SEMAG OR SEMG.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI;
RX   MEDLINE=96414430; PubMed=8817451;
RA   Adams R.H., Betz H., Pueschel A.W.;
RT   "A novel class of murine semaphorins with homology to thrombospondin
RT   is differentially expressed during early embryogenesis.";
RL   Mech. Dev. 57:33-45(1996).
CC   -!- FUNCTION: MAY ACT AS POSITIVE AXONAL GUIDANCE CUES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: IN ADULT, ONLY DETECTED IN BRAIN.
CC   -!- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED IN EMBRYONIC AND
CC       ADULT TISSUES. ITS ABUNDANCE DECREASES FROM E10 TO BITH.
CC   -!- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 SEMA DOMAIN.
CC   -!- SIMILARITY: CONTAINS 7 TSP TYPE-1 DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X97818; CAA66398.1; -.
DR   MGD; MGI:107555; Sema5b.
DR   InterPro; IPR003659; Plexin-like.
DR   InterPro; IPR002165; Plexin_repeat.
DR   InterPro; IPR001627; Sema.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 5.
DR   Pfam; PF01403; Sema; 1.
DR   Pfam; PF01437; PSI; 1.
DR   SMART; SM00423; PSI; 1.
DR   SMART; SM00209; TSP1; 4.
DR   PROSITE; PS50092; TSP1; 6.
KW   Signal; Transmembrane; Repeat; Multigene family; Neurogenesis;
KW   Developmental protein; Glycoprotein.
FT   SIGNAL        1     19       POTENTIAL.
FT   CHAIN        20   1093       SEMAPHORIN 5B.
FT   DOMAIN       20    978       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    979    999       POTENTIAL.
FT   DOMAIN     1000   1093       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      236    518       SEMA.
FT   DOMAIN      551    605       TSP TYPE-1 1.
FT   DOMAIN      606    663       TSP TYPE-1 2.
FT   DOMAIN      664    720       TSP TYPE-1 3.
FT   DOMAIN      721    776       TSP TYPE-1 4.
FT   DOMAIN      795    851       TSP TYPE-1 5.
FT   DOMAIN      852    908       TSP TYPE-1 6.
FT   DOMAIN      909    958       TSP TYPE-1 7.
FT   CARBOHYD     59     59       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     95     95       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    178    178       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    287    287       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    378    378       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    532    532       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    539    539       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    602    602       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    728    728       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    944    944       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1093 AA;  120326 MW;  29E5C9B1E8108717 CRC64;

  Query Match             22.7%;  Score 56.5;  DB 1;  Length 1093;
  Best Local Similarity   31.8%;  Pred. No. 16;
  Matches   14;  Conservative    7;  Mismatches   18;  Indels    5;  Gaps    2;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCP-YNGRLTTVKPRK 44
           | || : | | |:|    :  |    |:  | : ||:   | |:
Db     612 WSSWAQCSTSCGIGFQVRQRSC----SNPAPRHGGRICVGKSRE 651


RESULT 7
SSPO_BOVIN
ID   SSPO_BOVIN     STANDARD;      PRT;   867 AA.
AC   P98167;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   SCO-spondin (Fragment).
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ependymocyte;
RX   MEDLINE=96338614; PubMed=8743952;
RA   Gobron S., Monnerie H., Meiniel R., Creveaux I., Lehmann W.,
RA   Lamalle D., Dastugue B., Meiniel A.;
RT   "SCO-spondin: a new member of the thrombospondin family secreted by
RT   the subcommissural organ is a candidate in the modulation of neuronal
RT   aggregation.";
RL   J. Cell Sci. 109:1053-1061(1996).
CC   -!- FUNCTION: INVOLVED IN THE MODULATION OF NEURONAL AGGREGATION.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: SUBCOMMISSURAL ORGAN.
CC   -!- DEVELOPMENTAL STAGE: EMBRYO.
CC   -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
CC   -!- SIMILARITY: CONTAINS AT LEAST 4 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS AT LEAST 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS AT LEAST 1 F5/8 TYPE C DOMAIN.
CC   -!- SIMILARITY: CONTAINS AT LEAST 3 LDL-RECEPTOR CLASS A DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X93922; CAA63815.1; -.
DR   HSSP; P01130; 1AJJ.
DR   InterPro; IPR000421; FA58_C.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR002919; TIL_Cysrich.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00057; ldl_recept_a; 3.
DR   Pfam; PF00090; tsp_1; 4.
DR   Pfam; PF00093; vwc; 1.
DR   Pfam; PF00754; F5_F8_type_C; 1.
DR   Pfam; PF01826; TIL; 1.
DR   SMART; SM00231; FA58C; 1.
DR   SMART; SM00192; LDLa; 3.
DR   SMART; SM00209; TSP1; 4.
```

```
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01285; FA58C_1; 1.
DR   PROSITE; PS01286; FA58C_2; 1.
DR   PROSITE; PS01209; LDLRA_1; 3.
DR   PROSITE; PS50068; LDLRA_2; 3.
DR   PROSITE; PS50092; TSP1; 4.
KW   Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain.
FT   NON_TER       1      1
FT   DOMAIN       26     81       TSP TYPE-1 1.
FT   DOMAIN      103    142       EGF-LIKE 1.
FT   DOMAIN      143    180       EGF-LIKE 2.
FT   DOMAIN      185    243       TSP TYPE-1 2.
FT   DOMAIN      344    502       F5/8 TYPE C.
FT   DOMAIN      506    544       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN      663    701       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN      723    761       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN      762    813       TSP TYPE-1 3.
FT   DOMAIN      814    867       TSP TYPE-1 4.
FT   DISULFID    107    122       BY SIMILARITY.
FT   DISULFID    116    127       BY SIMILARITY.
FT   DISULFID    129    141       BY SIMILARITY.
FT   DISULFID    147    166       BY SIMILARITY.
FT   DISULFID    149    169       BY SIMILARITY.
FT   DISULFID    171    179       BY SIMILARITY.
FT   DISULFID    344    502       BY SIMILARITY.
FT   DISULFID    508    520       BY SIMILARITY.
FT   DISULFID    515    533       BY SIMILARITY.
FT   DISULFID    527    542       BY SIMILARITY.
FT   DISULFID    665    677       BY SIMILARITY.
FT   DISULFID    672    690       BY SIMILARITY.
FT   DISULFID    684    699       BY SIMILARITY.
FT   DISULFID    725    737       BY SIMILARITY.
FT   DISULFID    732    750       BY SIMILARITY.
FT   DISULFID    744    759       BY SIMILARITY.
FT   CARBOHYD     88     88       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    309    309       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    409    409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   NON_TER     867    867
SQ   SEQUENCE   867 AA;  91817 MW;  9538F2108E787B49 CRC64;

  Query Match             22.1%;  Score 55;  DB 1;  Length 867;
  Best Local Similarity   35.1%;  Pred. No. 20;
  Matches   13;  Conservative    4;  Mismatches   18;  Indels    2;  Gaps    1;

Qy       1 EWYSWKRSSKSTGLGDTATREGC--GPSQSDGCPYNG 35
           || :|   | | | |  : |  |   | :: | |  |
Db     821 EWEAWGPCSVSCGGGHRSRRRSCMDPPPKNGGAPCPG 857


RESULT 8
CSP_PLAFL
ID   CSP_PLAFL      STANDARD;      PRT;   315 AA.
AC   P05691;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Circumsporozoite protein (CS) (Fragment).
OS   Plasmodium falciparum (isolate le5).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5840;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87308186; PubMed=2442154;
RA   la Cruz V.F., Lal A.A., McCutchan T.F.;
RT   "Sequence variation in putative functional domains of the
RT   circumsporozoite protein of Plasmodium falciparum. Implications for
RT   vaccine development.";
RL   J. Biol. Chem. 262:11935-11939(1987).
CC   -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M17802; AAA29538.1; -.
KW   Malaria; Sporozoite; Repeat.
FT   NON_TER       1      1
FT   DOMAIN      107    270       40 X 4 AA TANDEM REPEATS OF P-N-A-N.
FT   NON_TER     315    315
SQ   SEQUENCE   315 AA;  33649 MW;  A334DB11FA7FD777 CRC64;

  Query Match             21.9%;  Score 54.5;  DB 1;  Length 315;
  Best Local Similarity   36.1%;  Pred. No. 8.1;
  Matches   13;  Conservative    4;  Mismatches   12;  Indels    7;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTAT-------REGCGPSQSDG 30
           ||| |::|:| |  |           |||    : ||
Db      46 WYSLKKNSRSLGENDDGNNNNGDNGREGKDEDKRDG 81


RESULT 9
CSP_PLAFA
ID   CSP_PLAFA      STANDARD;      PRT;   412 AA.
AC   P02893;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Circumsporozoite protein precursor (CS).
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=84250215; PubMed=6204383;
RA   Dame J.B., Williams J.L., McCutchan T.F., Weber J.L., Wirtz R.A.,
RA   Hochmeyer W.T., Maloy W.L., Haynes J.D., Schneider I., Roberts D.,
RA   Sanders G.S., Reddy E.P., Diggs C.L., Miller L.H.;
RT   "Structure of the gene encoding the immunodominant surface antigen on
RT   the sporozoite of the human malaria parasite Plasmodium falciparum.";
RL   Science 225:593-599(1984).
CC   -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; K02194; AAA29524.1; -.
DR   PIR; A03388; OZZQAF.
DR   InterPro; IPR003067; Crcmsprzoite.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PRINTS; PR01303; CRCMSPRZOITE.
DR   SMART; SM00209; TSP1; 1.
KW   Malaria; Sporozoite; Repeat; Signal.
FT   SIGNAL        1     16       PROBABLE.
```

```
FT   CHAIN        17    412       CIRCUMSPOROZOITE PROTEIN.
FT   DOMAIN      123    290       41 X 4 AA TANDEM REPEATS OF P-N-A-N.
SQ   SEQUENCE   412 AA;  44420 MW;  1EEEED3DE90965F8 CRC64;

  Query Match             21.9%;  Score 54.5;  DB 1;  Length 412;
  Best Local Similarity   36.1%;  Pred. No. 11;
  Matches   13;  Conservative    4;  Mismatches   12;  Indels    7;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTAT-------REGCGPSQSDG 30
           ||| |::|:| |  |           |||    : ||
Db      62 WYSLKKNSRSLGENDDGNNNNGDNGREGKDEDKRDG 97


RESULT 10
CSP_PLAFT
ID   CSP_PLAFT      STANDARD;      PRT;   424 AA.
AC   P13814;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Circumsporozoite protein precursor (CS).
OS   Plasmodium falciparum (isolate t4 / Thailand).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5846;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87315205; PubMed=3306373;
RA   del Portillo H.A., Nussenzweig R.S., Enea V.;
RT   "Circumsporozoite gene of a Plasmodium falciparum strain from
RT   Thailand.";
RL   Mol. Biochem. Parasitol. 24:289-294(1987).
CC   -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M19752; AAA29555.1; -.
DR   PIR; A54533; A54533.
DR   InterPro; IPR003067; Crcmsprzoite.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PRINTS; PR01303; CRCMSPRZOITE.
DR   SMART; SM00209; TSP1; 1.
KW   Malaria; Sporozoite; Repeat; Signal.
FT   SIGNAL        1     16       PROBABLE.
FT   CHAIN        17    424       CIRCUMSPOROZOITE PROTEIN.
FT   DOMAIN      123    300       45 X 4 AA TANDEM REPEATS OF N-A-N-P.
SQ   SEQUENCE   424 AA;  45610 MW;  710AB14238786CD9 CRC64;

  Query Match             21.9%;  Score 54.5;  DB 1;  Length 424;
  Best Local Similarity   36.1%;  Pred. No. 11;
  Matches   13;  Conservative    4;  Mismatches   12;  Indels    7;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTA-------TREGCGPSQSDG 30
           ||| |::|:| |  |           |||    : ||
Db      62 WYSLKKNSRSLGENDDGNNNNGDNNREGKDEDKRDG 97


RESULT 11
BAI1_HUMAN
ID   BAI1_HUMAN     STANDARD;      PRT;  1584 AA.
AC   O14514;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Brain-specific angiogenesis inhibitor 1 precursor.
GN   BAI1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98054121; PubMed=9393972;
RA   Nishimori H., Shiratsuchi T., Urano T., Kimura Y., Kiyono K.,
RA   Tatsumi K., Yoshida S., Ono M., Kuwano M., Nakamura Y., Tokino T.;
RT   "A novel brain-specific p53-target gene, BAI1, containing
RT   thrombospondin type 1 repeats inhibits experimental angiogenesis.";
RL   Oncogene 15:2145-2150(1997).
RN   [2]
RP   INTERACTION WITH BAP1.
RX   MEDLINE=98321173; PubMed=9647739;
RA   Shiratsuchi T., Futamura M., Oda K., Nishimori H., Nakamura Y.,
RA   Tokino T.;
RT   "Cloning and characterization of BAI-associated protein 1: a PDZ
RT   domain-containing protein that interacts with BAI1.";
RL   Biochem. Biophys. Res. Commun. 247:597-604(1998).
CC   -!- FUNCTION: LIKELY TO BE A POTENT INHIBITOR OF ANGIOGENESIS IN
CC       BRAIN AND MAY PLAY A SIGNIFICANT ROLE AS A MEDIATOR OF THE P53
CC       SIGNAL IN SUPPRESSION OF GLIOBLASTOMA. MAY FUNCTION IN CELL
CC       ADHESION AND SIGNAL TRANSDUCTION IN THE BRAIN.
CC   -!- SUBUNIT: INTERACTS WITH BAP1.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. LIKELY TO BE
CC       CONCENTRATED AT CELL-CELL ADHESION SITES.
CC   -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN BRAIN. REDUCED OR NO
CC       EXPRESSION IS OBSERVED IN SOME GLIOBLASTOMA CELL LINES AND CANCER
CC       TISSUES.
CC   -!- INDUCTION: BY P53.
CC   -!- DOMAIN: THE TSP1 REPEATS INHIBIT IN VIVO ANGIOGENESIS IN RAT
CC       CORNEA INDUCED BY BFGF.
CC   -!- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.
CC   -!- SIMILARITY: CONTAINS 5 TSP TYPE-1 DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB005297; BAA23647.1; -.
DR   Genew; HGNC:943; BAI1.
DR   MIM; 602682; -.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   InterPro; IPR000884; TSP1.
DR   InterPro; IPR001879; hormn_receptor.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00090; tsp_1; 5.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00209; TSP1; 5.
DR   PROSITE; PS50221; GPS; 1.
DR   PROSITE; PS00649; G_PROTEIN_RECEP_F2_1; FALSE_NEG.
DR   PROSITE; PS00650; G_PROTEIN_RECEP_F2_2; FALSE_NEG.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS50092; TSP1; 5.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein; Signal;
```

```
KW   Repeat; Cell adhesion.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1584       BRAIN-SPECIFIC ANGIOGENESIS INHIBITOR 1.
FT   DOMAIN       31    948       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    949    969       1 (POTENTIAL).
FT   DOMAIN      970    980       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    981   1001       2 (POTENTIAL).
FT   DOMAIN     1002   1008       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1009   1029       3 (POTENTIAL).
FT   DOMAIN     1030   1052       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1053   1073       4 (POTENTIAL).
FT   DOMAIN     1074   1093       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1094   1114       5 (POTENTIAL).
FT   DOMAIN     1115   1136       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1137   1157       6 (POTENTIAL).
FT   DOMAIN     1158   1166       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1167   1187       7 (POTENTIAL).
FT   DOMAIN     1188   1584       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      261    316       TSP TYPE-1 1.
FT   DOMAIN      354    408       TSP TYPE-1 2.
FT   DOMAIN      409    463       TSP TYPE-1 3.
FT   DOMAIN      467    521       TSP TYPE-1 4.
FT   DOMAIN      522    576       TSP TYPE-1 5.
FT   DOMAIN      881    938       GPS.
FT   DOMAIN     1411   1422       POLY-PRO.
FT   DOMAIN     1425   1430       POLY-PRO.
FT   SITE        231    233       CELL ATTACHMENT SITE (POTENTIAL).
FT   DOMAIN     1365   1584       NECESSARY FOR INTERACTION WITH BAP1.
FT   DOMAIN     1581   1584       INDISPENSABLE FOR INTERACTION WITH BAP1.
FT   CARBOHYD     64     64       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    401    401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    607    607       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    692    692       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    844    844       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    877    877       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    881    881       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1584 AA;  173531 MW;  DEA8F28C77874513 CRC64;

  Query Match           21.9%;  Score 54.5;  DB 1;  Length 1584;
  Best Local Similarity   30.2%;  Pred. No. 43;
  Matches   16;  Conservative    2;  Mismatches   18;  Indels   17;  Gaps    2;

Qy       1 EWYSWKRSSKSTGLG-DTATREGCGPSQSDG----------------CPYNGR 36
           || ||   | |   |    |||  |||                    || :|:
Db     472 EWSSWSACSASCSQGRQQRTRECNGPSYGGAECQGHWVETRDCFLQQCPVDGK 524


RESULT 12
GUX3_AGABI
ID   GUX3_AGABI     STANDARD;      PRT;   438 AA.
AC   P49075;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Exoglucanase 3 precursor (EC 3.2.1.91) (Exocellobiohydrolase 3)
DE   (1,4-beta-cellobiohydrolase 3).
GN   CEL3.
OS   Agaricus bisporus (Common mushroom).
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC   Agaricales; Agaricaceae; Agaricus.
OX   NCBI_TaxID=5341;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 255-277 AND 331-351.
RC   STRAIN=D649;
RX   MEDLINE=94368092; PubMed=8085821;
RA   Chow C.-M., Yague E., Raguz S., Wood D.A., Thurston C.F.;
RT   "The cel3 gene of Agaricus bisporus codes for a modular cellulase and
RT   is transcriptionally regulated by the carbon source.";
RL   Appl. Environ. Microbiol. 60:2779-2785(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mycelium;
RX   MEDLINE=96269930; PubMed=8662210;
RA   Yague E., Chow C.-M., Challen M.P., Thurston C.F.;
RT   "Correlation of exons with functional domains and folding regions in
RT   a cellulase from Agaricus bisporus.";
RL   Curr. Genet. 30:56-61(1996).
CC   -!- FUNCTION: SHOWS ENZYMATIC ACTIVITY TOWARDS CRYSTALLINE CELLULOSE.
CC       AT LONG REACTION TIMES. IT IS ALSO ABLE TO DEGRADE CARBOXYMETHYL
CC       CELLULOSE AND BARLEY B-GLUCAN.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC       in cellulose and cellotetraose, releasing cellobiose from the non-
CC       reducing ends of the chains.
CC   -!- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
CC       HYDROLASES).
CC   -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L24519; AAA50607.1; -.
DR   EMBL; L24520; AAA50608.1; -.
DR   EMBL; Z34007; CAA83971.1; -.
DR   HSSP; P00725; 1AZ6.
DR   InterPro; IPR000254; CBD_fungal.
DR   InterPro; IPR001524; GH_6.
DR   Pfam; PF00734; CBM_1; 1.
DR   Pfam; PF01341; Glyco_hydro_6; 1.
DR   PRINTS; PR00733; GLHYDRLASE6.
DR   ProDom; PD001821; CBD_fungal; 1.
DR   ProDom; PD003733; GH_6; 1.
DR   SMART; SM00236; fCBD; 1.
DR   PROSITE; PS00562; CBD_FUNGAL; 1.
DR   PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
DR   PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
KW   Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21    438       EXOGLUCANASE 3.
FT   DOMAIN       21     59       CELLULOSE-BINDING (BY SIMILARITY).
FT   DOMAIN       60     87       LINKER.
FT   DOMAIN       88    438       CATALYTIC.
FT   ACT_SITE    215    215       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    393    393       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     28     45       BY SIMILARITY.
FT   DISULFID     39     55       BY SIMILARITY.
FT   DISULFID    170    229       BY SIMILARITY.
FT   DISULFID    360    407       BY SIMILARITY.
FT   VARIANT     133    133       V -> T.
FT   VARIANT     152    152       R -> Q.
FT   VARIANT     244    244       V -> I.
FT   VARIANT     248    248       N -> D.
FT   VARIANT     398    398       N -> S.
SQ   SEQUENCE   438 AA;  46209 MW;  002C973544893794 CRC64;

  Query Match           21.7%;  Score 54;  DB 1;  Length 438;
  Best Local Similarity   38.5%;  Pred. No. 13;
  Matches   10;  Conservative    3;  Mismatches   13;  Indels    0;  Gaps    0;

Qy      17 TATREGCGPSQSDGCPYNGRLTTVKP 42
           | |  | ||:   | || |:   : |
Db      76 TTTSGGTGPTSGAGNPYTGKTVWLSP 101


RESULT 13
MPL1_LOLPR
ID   MPL1_LOLPR     STANDARD;      PRT;   263 AA.
AC   P14946; P19964;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Pollen allergen Lol p 1 precursor (Lol p I) (Allergen R7).
```

```
OS   Lolium perenne (Perennial ryegrass).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Poeae; Lolium.
OX   NCBI_TaxID=4522;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90375479; PubMed=1697854;
RA   Perez M., Ishioka G.Y., Walker L.E., Chesnut R.W.;
RT   "cDNA cloning and immunological characterization of the rye grass
RT   allergen Lol p I.";
RL   J. Biol. Chem. 265:16210-16215(1990).
RN   [2]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 24-48.
RC   TISSUE=Pollen;
RX   MEDLINE=91160716; PubMed=2001733;
RA   Griffith I.J., Smith P.M., Pollock J., Theerakulpisut P.,
RA   Avjoglu A., Davies S., Hough T., Singh M.B., Simpson R.J., Ward L.D.,
RA   Knox R.B.;
RT   "Cloning and sequencing of Lol pI, the major allergenic protein of
RT   rye-grass pollen.";
RL   FEBS Lett. 279:210-215(1991).
RN   [3]
RP   SEQUENCE OF 24-53.
RC   TISSUE=Pollen;
RX   MEDLINE=86242068; PubMed=3718469;
RA   Cottam G.P., Moran D.M., Standring R.;
RT   "Physicochemical and immunochemical characterization of allergenic
RT   proteins from rye-grass (Lolium perenne) pollen prepared by a rapid
RT   and efficient purification method.";
RL   Biochem. J. 234:305-310(1986).
RN   [4]
RP   SEQUENCE OF 236-263.
RX   MEDLINE=89364850; PubMed=2475768;
RA   Esch R.E., Klapper D.G.;
RT   "Isolation and characterization of a major cross-reactive grass group
RT   I allergenic determinant.";
RL   Mol. Immunol. 26:557-561(1989).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DISEASE: CAUSES GRASS POLLEN ALLERGY.
CC   -!- SIMILARITY: BELONGS TO THE EXPANSIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 EXPANSIN-LIKE EG45 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EXPANSIN-LIKE CBD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M57474; AAA63279.1; -.
DR   EMBL; M57476; AAA63278.1; -.
DR   PIR; A23341; A23341.
DR   PIR; B37881; B37881.
DR   PIR; S13614; S13614.
DR   HSSP; P43214; 1WHO.
DR   InterPro; IPR000882; Pollen_allergen.
DR   Pfam; PF01357; Pollen_allergen; 1.
DR   PRINTS; PR01225; EXPANSNFAMLY.
DR   ProDom; PD002179; Pollen_allergen; 1.
DR   PROSITE; PS50843; EXPANSIN_CBD; 1.
DR   PROSITE; PS50842; EXPANSIN_EG45; 1.
KW   Allergen; Glycoprotein; Signal; Multigene family.
FT   SIGNAL       1     23
FT   CHAIN       24    263       POLLEN ALLERGEN LOL P 1.
FT   DOMAIN      61    167       EXPANSIN-LIKE EG45.
FT   DOMAIN     181    262       EXPANSIN-LIKE CBD.
FT   CARBOHYD    32     32       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     68     68       N -> D.
FT   VARIANT    177    177       A -> G.
FT   VARIANT    210    210       I -> T.
FT   VARIANT    246    246       F -> V.
FT   CONFLICT    28     28       P -> C (IN REF. 3).
FT   CONFLICT    31     31       P -> C (IN REF. 3).
FT   CONFLICT    48     48       W -> WW (IN REF. 3).
SQ   SEQUENCE   263 AA;  28438 MW;  7675896F279C88C9 CRC64;

  Query Match             21.5%;  Score 53.5;  DB 1;  Length 263;
  Best Local Similarity   35.3%;  Pred. No. 9;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||       | || |        ||    |  |:||
Db      48 WY-----GKPTGAGPKDNGGACGYKNVDKAPFNG 76


RESULT 14
MPA1_PHAAQ
ID   MPA1_PHAAQ     STANDARD;      PRT;   269 AA.
AC   Q41260;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Major pollen allergen Pha a 1 precursor (Pha a I).
OS   Phalaris aquatica (Canary grass).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC   Aveneae; Phalaris.
OX   NCBI_TaxID=28479;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pollen;
RX   MEDLINE=96105569; PubMed=8564724;
RA   Suphioglu C., Singh M.B.;
RT   "Cloning, sequencing and expression in Escherichia coli of Pha a 1
RT   and four isoforms of Pha a 5, the major allergens of canary grass
RT   pollen.";
RL   Clin. Exp. Allergy 25:853-865(1995).
RN   [2]
RP   SEQUENCE OF 30-49.
RX   MEDLINE=93319091; PubMed=7687099;
RA   Suphioglu C., Singh M.B., Simpson R.J., Ward L., Knox R.B.;
RT   "Identification of canary grass (Phalaris aquatica) pollen allergens
RT   by immunoblotting: IgE and IgG antibody-binding studies.";
RL   Allergy 48:273-281(1993).
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DISEASE: CAUSES GRASS POLLEN ALLERGY.
CC   -!- SIMILARITY: BELONGS TO THE EXPANSIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 1 EXPANSIN-LIKE EG45 DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 EXPANSIN-LIKE CBD DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S80654; AAB35984.1; -.
DR   HSSP; P43214; 1WHO.
DR   InterPro; IPR000882; Pollen_allergen.
DR   Pfam; PF01357; Pollen_allergen; 1.
DR   PRINTS; PR01225; EXPANSNFAMLY.
DR   PROSITE; PS50843; EXPANSIN_CBD; 1.
DR   PROSITE; PS50842; EXPANSIN_EG45; 1.
KW   Allergen; Glycoprotein; Signal.
FT   SIGNAL       1     29
FT   CHAIN       30    269       MAJOR POLLEN ALLERGEN PHA A 1.
FT   DOMAIN      67    173       EXPANSIN-LIKE EG45.
FT   DOMAIN     187    268       EXPANSIN-LIKE CBD.
FT   CARBOHYD    38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    37     37       P -> G (IN REF. 2).
SQ   SEQUENCE   269 AA;  29011 MW;  05883A458ACE877F CRC64;
```

```
  Query Match             21.5%;  Score 53.5;  DB 1;  Length 269;
  Best Local Similarity   35.3%;  Pred. No. 9.2;
  Matches   12;  Conservative    1;  Mismatches   16;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNG 35
           ||      | || |       ||    |  |:||
Db      54 WY-----GKPTGAGPKDNGGACGYKDVDKAPFNG 82


RESULT 15
PILS_PSEAE
ID   PILS_PSEAE     STANDARD;      PRT;   530 AA.
AC   P33639;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sensor protein pilS (EC 2.7.3.-).
GN   PILS OR PA4546.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=93225810; PubMed=8097014;
RA   Hobbs M., Collie E.S.R., Free P.D., Livingston S.P., Mattick J.S.;
RT   "PilS and PilR, a two-component transcriptional regulatory system
RT   controlling expression of type 4 fimbriae in Pseudomonas
RT   aeruginosa.";
RL   Mol. Microbiol. 7:669-682(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM PILS/PILR
CC       INVOLVED IN THE REGULATION OF THE EXPRESSION OF TYPE 4 FIMBRIAE.
CC       PILS MAY FUNCTION AS A MEMBRANE-ASSOCIATED PROTEIN KINASE THAT
CC       PHOSPHORYLATES PILR IN RESPONSE TO ENVIRONMENTAL SIGNALS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: CONTAINS 1 HISTIDINE KINASE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z12154; CAA78138.1; -.
DR   EMBL; AE004868; AAG07934.1; -.
DR   PIR; S26600; S26600.
DR   PIR; S33673; S33673.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR004359; HIS_KIN_sig.
DR   InterPro; IPR003661; His_kinA.
DR   InterPro; IPR000014; PAS_domain.
DR   Pfam; PF00512; signal; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   SMART; SM00388; HisKA; 1.
DR   SMART; SM00091; PAS; 1.
DR   PROSITE; PS50109; HIS_KIN; 1.
DR   PROSITE; PS50112; PAS; FALSE_NEG.
KW   Sensory transduction; Transferase; Kinase; Phosphorylation;
KW   Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM     25     37       POTENTIAL.
FT   TRANSMEM     57     70       POTENTIAL.
FT   TRANSMEM     76     98       POTENTIAL.
FT   TRANSMEM    101    119       POTENTIAL.
FT   TRANSMEM    124    144       POTENTIAL.
FT   TRANSMEM    156    174       POTENTIAL.
FT   DOMAIN      175    530       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      196    260       PAS.
FT   DOMAIN      316    527       HISTIDINE KINASE.
FT   MOD_RES     319    319       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   530 AA;  58985 MW;  1F6E0300E2E77A4C CRC64;

  Query Match             21.5%;  Score 53.5;  DB 1;  Length 530;
  Best Local Similarity   37.0%;  Pred. No. 19;
  Matches   17;  Conservative    4;  Mismatches   12;  Indels   13;  Gaps    2;

Qy       9 SKSTGLGDTATREGCGPSQS----------DGCPYNGRLTTVKPRK 44
           || ||||   :|| |  :|:           ||    |:|   |||
Db     486 SKGTGLGLYLSRELCESNQARIDYRNREEGGGC---FRITFAHPRK 528


Search completed: January 15, 2003, 08:15:52
Job time : 3.95434 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:27 ; Search time 2.96347 Seconds
(without alignments)
3059.276 Million cell updates/sec

Title: US-09-079-819-50
Perfect score: 249
Sequence: 1 EWYSWKRSSKSTGLGDTATR.....PSQSDGCPYNGRLTTVKPRK 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	68	27.3	793	5	Q9VNH0	Q9vnh0 drosophila
2	64.5	25.9	415	5	O44228	O44228 halocynthia
3	64	25.7	503	11	Q9ERZ0	Q9erz0 mus musculu
4	63	25.3	109	2	Q9X2S0	Q9x2s0 marine prot
5	63	25.3	109	2	Q9F8Y4	Q9f8y4 marine stro
6	63	25.3	485	2	Q53737	Q53737 streptomyce
7	62.5	25.1	316	10	Q9LI19	Q9li19 oryza sativ
8	60	24.1	136	2	Q93P23	Q93p23 anaplasma m
9	59.5	23.9	679	2	Q93AS9	Q93as9 ralstonia s
10	59	23.7	275	5	Q9VLX6	Q9vlx6 drosophila
11	59	23.7	304	2	Q57075	Q57075 s integron
12	59	23.7	503	16	Q9RK46	Q9rk46 streptomyce
13	59	23.7	514	2	Q9WTH9	Q9wth9 escherichia
14	59	23.7	682	12	Q8QS63	Q8qs63 chimpanzee
15	59	23.7	1360	5	Q9TYK4	Q9tyk4 caenorhabdi
16	58.5	23.5	2533	5	Q27183	Q27183 paramecium
17	58.5	23.5	2533	5	P90589	P90589 paramecium
18	58	23.3	109	2	Q8VPH9	Q8vph9 uncultured
19	58	23.3	727	11	Q8R0X1	Q8r0x1 mus musculu
20	57.5	23.1	509	16	Q8ZP41	Q8zp41 salmonella
21	57.5	23.1	509	16	Q8Z7F5	Q8z7f5 salmonella
22	57.5	23.1	4123	4	O75851	O75851 homo sapien
23	57	22.9	108	2	O54006	O54006 paenibacill
24	57	22.9	108	2	O54010	O54010 paenibacill
25	57	22.9	140	2	Q8VSI3	Q8vsi3 shigella fl
26	57	22.9	476	11	Q8R1H9	Q8r1h9 mus musculu
27	57	22.9	657	11	Q8R0K8	Q8r0k8 mus musculu
28	57	22.9	735	4	O95784	O95784 homo sapien
29	56.5	22.7	332	12	Q82493	Q82493 influenza a
30	56.5	22.7	2673	4	Q96SC3	Q96sc3 homo sapien
31	56.5	22.7	5636	4	Q96RW7	Q96rw7 homo sapien
32	56	22.5	110	17	Q9YD85	Q9yd85 aeropyrum p
33	56	22.5	222	11	Q8VDZ0	Q8vdz0 mus musculu
34	56	22.5	234	2	O85452	O85452 rhizobium l
35	56	22.5	301	16	P95073	P95073 mycobacteri
36	56	22.5	459	11	Q9D1P1	Q9d1p1 mus musculu
37	56	22.5	472	10	Q9FIU1	Q9fiu1 arabidopsis
38	56	22.5	921	5	Q969A3	Q969a3 branchiosto
39	56	22.5	1444	5	Q17591	Q17591 caenorhabdi
40	56	22.5	1513	3	Q9HDV4	Q9hdv4 schizosacch
41	56	22.5	1705	5	Q9V4I8	Q9v4i8 drosophila
42	55.5	22.3	104	5	Q24949	Q24949 fasciola he
43	55.5	22.3	273	5	Q8SVB2	Q8svb2 encephalito
44	55.5	22.3	292	3	Q06896	Q06896 saccharomyc
45	55.5	22.3	342	5	Q16920	Q16920 aedes aegyp

ALIGNMENTS

RESULT 1
Q9VNH0
ID Q9VNH0 PRELIMINARY; PRT; 793 AA.
AC Q9VNH0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE CG2102 protein (LD36057P).
GN CAS OR CG2102.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

```
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003602; AAF51966.1; -.
DR   EMBL; AY051865; AAK93289.1; -.
DR   FlyBase; FBgn0004878; cas.
DR   InterPro; IPR000637; AT_hook.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR000822; Znf_C2H2.
DR   Pfam; PF02178; AT_hook; 1.
DR   Pfam; PF00096; zf-C2H2; 3.
DR   SMART; SM00384; AT_hook; 1.
DR   SMART; SM00355; ZnF_C2H2; 3.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 2.
KW   DNA-binding; Metal-binding; Zinc-finger.
SQ   SEQUENCE   793 AA;  87925 MW;  E35DE18DD3C37671 CRC64;

  Query Match             27.3%;  Score 68;  DB 5;  Length 793;
  Best Local Similarity   26.3%;  Pred. No. 1.5;
  Matches   10;  Conservative   10;  Mismatches   18;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLT 38
           :|:  :: |   |    :: : | |:  :|| || : |
Db     400 KWHKKRKESLKLGFARFSSSDDCAPAYGEGCAYNWKQT 437


RESULT 2
O44228
ID   O44228      PRELIMINARY;      PRT;   415 AA.
AC   O44228;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   HrTT-1 protein.
GN   HRTT-1.
OS   Halocynthia roretzi (Sea squirt).
OC   Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC   Stolidobranchia; Pyuridae; Halocynthia.
OX   NCBI_TaxID=7729;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98267369; PubMed=9601991;
RA   Hotta K., Takahashi H., Satoh N.;
RT   "Expression of an ascidian gene in the tip of the tail of tail-bud-
RT   stage embryos.";
RL   Dev. Genes Evol. 208:164-167(1998).
DR   EMBL; AB009609; BAA24011.1; -.
DR   HSSP; P04284; 1CFE.
DR   InterPro; IPR001283; Allrgn_V5/Tpx1.
DR   InterPro; IPR000884; TSP1.
DR   Pfam; PF00188; SCP; 1.
DR   Pfam; PF00090; tsp_1; 1.
DR   PRINTS; PR00837; V5TPXLIKE.
DR   ProDom; PD000542; Allrgn_V5/Tpx1; 1.
DR   SMART; SM00198; SCP; 1.
DR   SMART; SM00209; TSP1; 1.
DR   PROSITE; PS01009; SCP_AG5_PR1_SC7_1; 1.
DR   PROSITE; PS01010; SCP_AG5_PR1_SC7_2; 1.
DR   PROSITE; PS50092; TSP1; 1.
SQ   SEQUENCE   415 AA;  47365 MW;  22366EB3F19DEDD3 CRC64;

  Query Match             25.9%;  Score 64.5;  DB 5;  Length 415;
  Best Local Similarity   31.9%;  Pred. No. 2.1;
  Matches   15;  Conservative    3;  Mismatches   26;  Indels    3;  Gaps    1;

Qy       1 EWYSWKRSSKSTGLGDTATREGCG---PSQSDGCPYNGRLTTVKPRK 44
           || :|   ||| |:| |      |    |       |   :    || |
Db     319 EWTTWSSCSKSCGVGSTRRERQCNTFVPGDCKDFPSEVKFCVKKPCK 365


RESULT 3
Q9ERZ0
ID   Q9ERZ0      PRELIMINARY;      PRT;   503 AA.
AC   Q9ERZ0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hemogen.
GN   HGN.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RX   MEDLINE=21297709; PubMed=11404085;
RA   Yang L.V., Nicholson R.H., Kaplan J., Galy A., Li L.;
RT   "Hemogen is a novel nuclear factor specifically expressed in mouse
RT   hematopoietic development and its human homologue EDAG maps to
RT   chromosome 9q22, a region containing breakpoints of hematological
RT   neoplasms.";
RL   Mech. Dev. 104:105-111(2001).
DR   EMBL; AF269248; AAG33367.1; -.
DR   MGD; MGI:2136910; Hgn.
SQ   SEQUENCE   503 AA;  55043 MW;  B1027B81B55BB08D CRC64;

  Query Match             25.7%;  Score 64;  DB 11;  Length 503;
  Best Local Similarity   44.4%;  Pred. No. 3;
  Matches   16;  Conservative    3;  Mismatches   11;  Indels    6;  Gaps    2;

Qy      15 GDTATREGCGPS---QSD---GCPYNGRLTTVKPRK 44
           || |  ||| |    |||   ||| :   |:| | :
Db     288 GDAAGLEGCAPKALPQSDVAEGCPLDTTPTSVTPEQ 323


RESULT 4
Q9X2S0
ID   Q9X2S0      PRELIMINARY;      PRT;   109 AA.
AC   Q9X2S0;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Nitrogenase reductase (Fragment).
GN   NIFH.
OS   marine proteobacterium 'Tomales Bay med white'.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae.
OX   NCBI_TaxID=75784;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TOMALES BAY MED WHITE;
```

```
RA   Olson J.B., Litaker R.W., Paerl H.W.;
RT   "Ubiquity of heterotrophic diazotrophs in marine microbial mats.";
RL   Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF046852; AAD23124.1; -.
DR   HSSP; P00459; 1N2C.
DR   InterPro; IPR000392; NitrogenaseII.
DR   Pfam; PF00142; fer4_NifH; 1.
DR   PRINTS; PR00091; NITROGNASEII.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
FT   NON_TER       1      1
FT   NON_TER     109    109
SQ   SEQUENCE   109 AA;  11736 MW;  4134F2865F3BC73E CRC64;

  Query Match             25.3%;  Score 63;  DB 2;  Length 109;
  Best Local Similarity   41.4%;  Pred. No. 0.78;
  Matches   12;  Conservative    2;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      12 TGLGDTATREGCGPSQSDGCPYNGRLTTV 40
           || ||    |  ||   |||   | :| :
Db      35 TGYGDVKCAESGGPEPGDGCAGRGVITAI 63


RESULT 5
Q9F8Y4
ID   Q9F8Y4      PRELIMINARY;      PRT;   109 AA.
AC   Q9F8Y4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Dinitrogenase reductase NifH (Fragment).
GN   NIFH.
OS   marine stromatolite eubacterium HB(0898) Z07.
OC   Bacteria; environmental samples.
OX   NCBI_TaxID=137314;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Steppe T.F., Pinckney J.L., Dyble J., Paerl H.W.;
RT   "Diazotrophy in modern marine Bahamian stromatolites.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF227938; AAG23905.1; -.
DR   HSSP; P00456; 1CP2.
DR   InterPro; IPR000392; NitrogenaseII.
DR   Pfam; PF00142; fer4_NifH; 1.
DR   PRINTS; PR00091; NITROGNASEII.
DR   PROSITE; PS00746; NIFH_FRXC_1; 1.
DR   PROSITE; PS00692; NIFH_FRXC_2; 1.
FT   NON_TER       1      1
FT   NON_TER     109    109
SQ   SEQUENCE   109 AA;  11612 MW;  85024C123F0A101D CRC64;

  Query Match             25.3%;  Score 63;  DB 2;  Length 109;
  Best Local Similarity   38.7%;  Pred. No. 0.78;
  Matches   12;  Conservative    4;  Mismatches   15;  Indels    0;  Gaps    0;

Qy      10 KSTGLGDTATREGCGPSQSDGCPYNGRLTTV 40
           : || |||   |  ||    ||   | :|::
Db      31 RKTGFGDTLCTESGGPEPGVGCAGRGIITSI 61


RESULT 6
Q53737
ID   Q53737      PRELIMINARY;      PRT;   485 AA.
AC   Q53737;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   N-acetylpuromycin N-acetylhydrolase precursor.
GN   NAPH.
OS   Streptomyces lipmanii (Streptomyces alboniger).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC12461;
RX   MEDLINE=96139493; PubMed=8576156;
RA   Tercero J.A., Espinosa J.C., Lacalle R.A., Jimenez A.;
RT   "The biosynthetic pathway of the aminonucleoside antibiotic Puromycin,
RT   as deduced from the molecular analysis of the pur cluster of
RT   Streptomyces alboniger.";
RL   J. Biol. Chem. 271:1579-1590(1996).
DR   EMBL; X92429; CAA63158.1; -.
DR   HSSP; P80561; 1XJO.
DR   InterPro; IPR003137; PA.
DR   Pfam; PF02225; PA; 1.
KW   Hydrolase; Signal.
FT   SIGNAL        1     33      POTENTIAL.
SQ   SEQUENCE   485 AA;  51358 MW;  15016CCF70E5EBB0 CRC64;

  Query Match             25.3%;  Score 63;  DB 2;  Length 485;
  Best Local Similarity   46.2%;  Pred. No. 4;
  Matches   12;  Conservative    2;  Mismatches   12;  Indels    0;  Gaps    0;

Qy      15 GDTATREGCGPSQSDGCPYNGRLTTV 40
           |||  |:||     || |  ||:  |
Db     141 GDTERRQGCRAGAYDGVPAGGRIALV 166


RESULT 7
Q9LI19
ID   Q9LI19      PRELIMINARY;      PRT;   316 AA.
AC   Q9LI19;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical protein.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NIPPONBARE;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0708G02.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP001539; BAA92919.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   316 AA;  33263 MW;  7D2BFA7A90C9C8B0 CRC64;

  Query Match             25.1%;  Score 62.5;  DB 10;  Length 316;
  Best Local Similarity   31.1%;  Pred. No. 2.9;
  Matches   14;  Conservative    9;  Mismatches   17;  Indels    5;  Gaps    1;

Qy       4 SWKRSSKSTGLG-----DTATREGCGPSQSDGCPYNGRLTTVKPR 43
           ||: |:  :| |     |     ||| : :|    :  ||:::||
Db      54 SWRSSNSHSGEGGVSHDDDEAGSGCGEADTDENSSDVNLTSLRPR 98


RESULT 8
Q93P23
ID   Q93P23      PRELIMINARY;      PRT;   136 AA.
AC   Q93P23;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Major surface protein 2 hypervariable region (Fragment).
OS   Anaplasma marginale.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Anaplasmataceae; Anaplasma.
OX   NCBI_TaxID=770;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=VIRGINIA;
RX   MEDLINE=21340406; PubMed=11447197;
RA   de La Fuente J., Kocan K.M.;
RT   "Expression of Anaplasma marginale Major Surface Protein 2 Variants in
RT   Persistently Infected Ticks.";
RL   Infect. Immun. 69:5151-5156(2001).
DR   EMBL; AF354475; AAK76937.1; -.
DR   InterPro; IPR002566; Surface_Ag_msp4.
DR   Pfam; PF01617; Surface_Ag_2; 1.
FT   NON_TER       1       1
FT   NON_TER     136     136
SQ   SEQUENCE   136 AA;  13782 MW;  FEEEFA8BB59DD661 CRC64;

  Query Match             24.1%;  Score 60;  DB 2;  Length 136;
  Best Local Similarity   39.4%;  Pred. No. 2.5;
  Matches   13;  Conservative    5;  Mismatches   15;  Indels    0;  Gaps    0;

Qy       8 SSKSTGLGDTATREGCGPSQSDGCPYNGRLTTV 40
           | |  | | |:    || : : |   ||:|:||
Db      27 SKKVCGKGTTSGSNQCGVNATSGSTNNGKLSTV 59


RESULT 9
Q93AS9
ID   Q93AS9      PRELIMINARY;      PRT;   679 AA.
AC   Q93AS9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   PehC.
GN   PEHC.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K60;
RA   Gonzalez E.T., Allen C.;
RT   "Identifying the role of pectolysis during Ralstonia solanacearum
RT   pathogenesis on tomato.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
CC       (POLYGALACTURONASES).
DR   EMBL; AF417111; AAL24033.1; -.
DR   InterPro; IPR000743; GH28.
DR   Pfam; PF00295; Glyco_hydro_28; 1.
DR   PROSITE; PS00502; POLYGALACTURONASE; UNKNOWN_1.
KW   Cell wall; Glycosidase; Hydrolase.
SQ   SEQUENCE   679 AA;  70356 MW;  2033B13506E5AA66 CRC64;

  Query Match             23.9%;  Score 59.5;  DB 2;  Length 679;
  Best Local Similarity   34.1%;  Pred. No. 17;
  Matches   14;  Conservative    8;  Mismatches   14;  Indels    5;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKP 42
           |:::| :  : |: | :|     |||: | | |  | |  |
Db     268 WWTFKGTKGAYGVVDAST-----PSQASGNPNNVDLRTAAP 303


RESULT 10
Q9VLX6
ID   Q9VLX6      PRELIMINARY;      PRT;   275 AA.
AC   Q9VLX6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   CG7191 protein.
GN   CG7191.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003618; AAF52553.1; -.
DR   FlyBase; FBgn0031945; CG7191.
SQ   SEQUENCE   275 AA;  30277 MW;  97339BFD35FCB883 CRC64;

  Query Match             23.7%;  Score 59;  DB 5;  Length 275;
  Best Local Similarity   40.7%;  Pred. No. 7.3;
  Matches   11;  Conservative    4;  Mismatches   12;  Indels    0;  Gaps    0;

Qy       3 YSWKRSSKSTGLGDTATREGCGPSQSD 29
           | ||| :|: |  :   |  | ||: |
Db     209 YPWKRHNKNRGNYEPELRYACTPSKED 235


RESULT 11
Q57075
ID   Q57075      PRELIMINARY;      PRT;   304 AA.
AC   Q57075;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   INTEGRON IN2 FOUND in TN21 (SHIGELLA FLEXNERI) SULPHONAMIDE resistant
DE   dihydropteroate synthase (SUL1), ISTB (ISTB), ISTA (ISTA), TNIBDELTA1
DE   (TNIBDELTA1) and TNIA (TNIA) genes, complete CDS (SHIGELLA FLEXNERI)
DE   (ISTB) (TNIBDELTA1) (TNIA).
OS   Shigella flexneri,
OS   unidentified, and
OS   Escherichia coli.
OG   Plasmid IncFII R100.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
```

```
OX   NCBI_TaxID=623, 32644, 562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=TN21;
RA   Brown H.J., Stokes H.W., Hall R.M.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=TN21;
RA   Hall R.M.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=TN21;
RA   Brown H.J., Stokes H.W., Hall R.M.;
RL   Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96345606; PubMed=8755869;
RA   Brown H.J., Stokes H.W., Hall R.M.;
RT   "The integrons In0, In2, and In5 are defective transposon
RT   derivatives.";
RL   J. Bacteriol. 178:4429-4437(1996).
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Liebert C.A., Hall R.M., Summers A.O.;
RT   "Transposon Tn21: flagship of the floating genome.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U42226; AAC53731.1; -.
DR   EMBL; U40482; AAC44292.1; -.
DR   EMBL; AF071413; AAC33919.1; -.
DR   InterPro; IPR001584; Rve.
DR   Pfam; PF00665; rve; 1.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   304 AA;  35141 MW;  8F5AB489F3C8A784 CRC64;

  Query Match           23.7%;  Score 59;  DB 2;  Length 304;
  Best Local Similarity   29.5%;  Pred. No. 8.2;
  Matches   13;  Conservative   11;  Mismatches   18;  Indels    2;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTT--VKPR 43
           | :| | :::  |  : :::|| |  :    : ||| |  : ||
Db     214 WPAWIRRTENAQLTRSMSKKGCSPDNAACEGFFGRLKTELIYPR 257


RESULT 12
Q9RK46
ID   Q9RK46      PRELIMINARY;      PRT;   503 AA.
AC   Q9RK46;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative dioxygenase.
GN   SCO0333 OR SCF12.12C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Murphy L., Harris D.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL117669; CAB56138.1; -.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   InterPro; IPR004294; RPE65.
DR   Pfam; PF03055; RPE65; 1.
DR   PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW   Dioxygenase.
SQ   SEQUENCE   503 AA;  55189 MW;  EE3335F6EDCDCC50 CRC64;

  Query Match           23.7%;  Score 59;  DB 16;  Length 503;
  Best Local Similarity   36.7%;  Pred. No. 14;
  Matches   18;  Conservative    6;  Mismatches   15;  Indels   10;  Gaps    4;

Qy       1 EWY--SWKRSSKSTG-LGDTATREGCGPSQSDGC-----PYNGRLTTVK 41
           |||   | |||: |  ||:  |  |  |     |     ||:||:  ::
Db     119 EWYRNRWVRSSQVTAKLGE--TYPGQAPPDDFPCNTHVIPYHGRILALQ 165


RESULT 13
Q9WTH9
ID   Q9WTH9      PRELIMINARY;      PRT;   514 AA.
AC   Q9WTH9;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Possible transposase of IS1353.
GN   YAHA.
OS   Escherichia coli.
OG   Plasmid R100.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sampei G., Mizobuchi K.;
RT   "Organization and diversification of plasmid genomes: complete
RT   nucleotide sequence of the R100 genome.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000342; BAA78795.1; -.
DR   InterPro; IPR001584; Rve.
DR   InterPro; IPR002514; Transposase_8.
DR   Pfam; PF00665; rve; 1.
DR   Pfam; PF01527; Transposase_8; 1.
SQ   SEQUENCE   514 AA;  59587 MW;  9984076FB09FE9CA CRC64;

  Query Match           23.7%;  Score 59;  DB 2;  Length 514;
  Best Local Similarity   29.5%;  Pred. No. 14;
  Matches   13;  Conservative   11;  Mismatches   18;  Indels    2;  Gaps    1;

Qy       2 WYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLTT--VKPR 43
           | :| | :::  |  : :::|| |  :    : ||| |  : ||
Db     424 WPAWIRRTENAQLTRSMSKKGCSPDNAACEGFFGRLKTELIYPR 467


RESULT 14
```

```
Q8QS63
ID   Q8QS63      PRELIMINARY;      PRT;   682 AA.
AC   Q8QS63;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Tegument protein UL25.
OS   chimpanzee cytomegalovirus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Betaherpesvirinae; Cytomegalovirus.
OX   NCBI_TaxID=188763;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Davison A.J., Akter P., Dolan A., Wright K.M., Addison C.,
RA   Alcendor D.J., Hayward G.S., McGeoch D.J.;
RT   "The human cytomegalovirus genome revisited.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF480884; AAM00675.1; -.
SQ   SEQUENCE   682 AA;  75470 MW;  94148CD4C7A584DD CRC64;

  Query Match             23.7%;  Score 59;  DB 12;  Length 682;
  Best Local Similarity   35.1%;  Pred. No. 20;
  Matches   13;  Conservative    5;  Mismatches   19;  Indels    0;  Gaps    0;

Qy       8 SSKSTGLGDTATREGCGPSQSDGCPYNGRLTTVKPRK 44
           |   | |  ::|  | ||::  |    || |  || :
Db      81 SVSGTSLSSSSTPRGAGPNRGGGETGGGRTTAGKPSR 117


RESULT 15
Q9TYK4
ID   Q9TYK4      PRELIMINARY;      PRT;  1360 AA.
AC   Q9TYK4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 144.3 kDa protein.
GN   Y8A9A.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Courtney L., Langston Y., Drone K., Mead K.;
RT   "The sequence of C. elegans cosmid Y8A9A.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF125461; AAK18995.1; -.
DR   InterPro; IPR000884; TSP1.
DR   SMART; SM00209; TSP1; 5.
DR   PROSITE; PS50092; TSP1; 7.
KW   Hypothetical protein.
SQ   SEQUENCE   1360 AA;  144314 MW;  312B206B6D087C3A CRC64;

  Query Match             23.7%;  Score 59;  DB 5;  Length 1360;
  Best Local Similarity   36.8%;  Pred. No. 42;
  Matches   14;  Conservative    3;  Mismatches   21;  Indels    0;  Gaps    0;

Qy       1 EWYSWKRSSKSTGLGDTATREGCGPSQSDGCPYNGRLT 38
           :| :|   | : |   | ||     |   |||  | ||
Db    1207 DWGAWTTCSATCGGCGTQTRSRTCASAPYGCPCTGDLT 1244


Search completed: January 15, 2003, 08:19:25
Job time : 5.96347 secs
```

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:24 ; Search time 62.6849 Seconds
(without alignments)
1768.599 Million cell updates/sec

Title: US-09-079-819-178
Perfect score: 4321
Sequence: 1 MILQAHLHSLCLLMLYLATG.....DKGKDNVESAQASEVKPLRS 832

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4321	100.0	832	20	AAW74089	Human HPT-1 protei
2	4321	100.0	832	22	AAM24522	C880P similar amin
3	908.5	21.0	807	21	AAY99405	Human PRO1340 (UNQ
4	908.5	21.0	807	22	AAB87574	Human PRO1340. Ho
5	908.5	21.0	807	22	AAB66154	Protein of the inv
6	908.5	21.0	807	23	AAU83685	Human PRO protein,
7	877.5	20.3	830	22	AAU09959	Human cadherin-2 (
8	699.5	16.2	906	21	AAY70741	Human N-cadherin.
9	699.5	16.2	906	22	ABG14316	Novel human diagno
10	699.5	16.2	906	23	ABB81474	Human N-cadherin p
11	696	16.1	912	23	ABB81471	Chicken N-cadherin
12	694.5	16.1	906	23	ABB57233	Mouse ischaemic co
13	684.5	15.8	916	18	AAW25658	Human cadherin-4.
14	684.5	15.8	916	18	AAW13129	Full length human
15	671.5	15.5	896	15	AAR63533	Human HT-1376 cell
16	671.5	15.5	896	23	AAU78055	Human desmocollin
17	667.5	15.4	847	22	AAM39235	Human polypeptide
18	665.5	15.4	862	22	AAM41021	Human polypeptide
19	657.5	15.2	713	18	AAW25638	Human cadherin-13.
20	657.5	15.2	713	18	AAW13136	Full length human
21	630.5	14.6	712	13	AAR27823	Sequence encoded b
22	630.5	14.6	717	13	AAR27824	Sequence encoded b
23	611	14.1	4555	23	AAM52106	Rat fat 3 protein
24	607	14.0	759	22	ABG21947	Novel human diagno
25	605	14.0	4591	22	ABG22977	Novel human diagno
26	604	14.0	899	23	ABB81472	Mouse E-cadherin p
27	601.5	13.9	3606	22	ABB62595	Drosophila melanog
28	593	13.7	878	15	AAR55060	Sequence of human
29	593	13.7	878	16	AAR85487	Human E-cadherin p
30	593	13.7	878	23	ABB81475	Human E-cadherin p
31	592.5	13.7	117	21	AAY64597	Nonclassical cadhe
32	591.5	13.7	4643	22	ABB71609	Drosophila melanog
33	591	13.7	882	20	AAY09375	Wild-type E-cadher
34	591	13.7	882	21	AAB35730	Human E-cadherin a
35	591	13.7	882	22	AAB73490	Human E-cadherin.
36	591	13.7	882	23	AAU78051	Human E-cadherin p
37	576	13.3	3014	22	AAU68533	Human novel cytoki
38	576	13.3	3014	22	AAU02196	Seven-pass transme
39	576	13.3	3028	22	AAE08586	Human NOV7 protein
40	574.5	13.3	796	22	AAB60418	Mouse PrP-binding
41	573.5	13.3	796	15	AAR49730	Sequence encoded b
42	571.5	13.2	788	22	AAM78375	Human protein SEQ
43	571.5	13.2	796	23	ABB57294	Mouse ischaemic co
44	570.5	13.2	840	22	ABG01693	Novel human diagno
45	569.5	13.2	840	22	AAM23903	Human EST encoded

ALIGNMENTS

RESULT 1
AAW74089
ID AAW74089 standard; Protein; 832 AA.
XX
AC AAW74089;
XX
DT 04-MAY-1999 (first entry)
XX
DE Human HPT-1 protein sequence.
XX
KW Gastro-intestinal transport receptor; binding protein; hSI; HPT1;
KW D2H; hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;
KW intestinal peptide-associated transporter; hypertension; diabetes;
KW osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;
KW therapeutic agent delivery; therapy; probe.
XX
OS Homo sapiens.
XX
PN WO9851325-A2.
XX
PD 19-NOV-1998.
XX
PF 15-MAY-1998; 98WO-US10088.
XX
PR 15-MAY-1997; 97US-0046595.
XX
PA (CYTO-) CYTOGEN CORP.
PA (ELAN-) ELAN CORP PLC.
XX
PI Alvarez VL, Belinka BA, Cagney GM, Carter JM, Lambkin IJ;
PI Omahony DJ, Patterson CA, Singleton J;
XX
DR WPI; 1999-009568/01.

```
DR   N-PSDB; AAX18166.
XX
PT   New proteins that bind specifically to receptors in the
PT   gastro-intestinal tract and related nucleic acid - chimaeras and
PT   antibodies, used to deliver therapeutic or diagnostic agents to, or
PT   through, the gastrointestinal tract, e.g. insulin or leuprolide
XX
PS   Disclosure; Fig 2; 294pp; English.
XX
CC   This sequence is the human HPT-1 protein. The invention relates to
CC   purified proteins (I) that bind specifically to at least one of the
CC   gastro-intestinal (GI) tract receptors human intestinal
CC   peptide-associated transporter (HPT1), hPEPT1, D2H and human
CC   sucrose-isomaltase complex (hSI). (I) provide active transport of
CC   therapeutic agents through human and animal GI tissue (into the blood)
CC   for in vivo delivery, particularly for treatment or prevention
CC   of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,
CC   migraine, or angina pectoris. Specifically they are used to deliver
CC   insulin or leuprolide, but many other suitable therapeutic agents are
CC   disclosed, including genes or inhibitory nucleic acid, imaging agents and
CC   antigens. (I) may also provide targeting to the GI tract. Other uses of
CC   (I) are: (i) to determine the level of specified receptors in a sample
CC   (in a binding assay); and (ii) to screen for molecules that bind (I).
CC   Immunogenic analogues or derivatives of (I) are used to raise antibodies
CC   and in immunoassays. The antibodies are used to locate, detect and
CC   measure (I), e.g. for imaging, monitoring treatment, tissue analysis
CC   etc., also for peptide purification and immobilisation.
XX
SQ   Sequence   832 AA;

  Query Match            100.0%;  Score 4321;  DB 20;  Length 832;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches  832;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832


RESULT 2
AAM24522
ID   AAM24522 standard; Protein; 832 AA.
XX
AC   AAM24522;
XX
DT   12-OCT-2001  (first entry)
XX
DE   C880P similar amino acid sequence (GENESEQ W740898).
XX
KW   Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW   immunogenic; gene therapy; vaccine; colonic cancer.
XX
OS   Homo sapiens.
XX
PN   WO200149716-A2.
XX
PD   12-JUL-2001.
XX
PF   29-DEC-2000; 2000WO-US35596.
XX
PR   30-DEC-1999;   99US-0476296.
PR   10-JAN-2000; 2000US-0480321.
PR   15-FEB-2000; 2000US-0504629.
PR   06-MAR-2000; 2000US-0519444.
PR   19-MAY-2000; 2000US-0575251.
PR   29-JUN-2000; 2000US-0609448.
PR   28-AUG-2000; 2000US-0649811.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Lodes MJ,  Secrist H,  Benson DR,  Meagher MJ,  Stolk JA;
PI   King GE,  Wang T,  Jiang Y;
XX
DR   WPI; 2001-441847/47.
XX
PT   Colon tumor associated proteins and nucleic acids useful for the
PT   prevention, diagnosis and treatment of colonic cancer -
XX
PS   Claim 2; Page 469-472; 472pp; English.
XX
CC   The present invention describes colon tumour associated proteins (I) and
CC   the polynucleotides (II) that encode them. (I) have cytostatic activity.
CC   (I) and (II) can be used in gene therapy and vaccine production. (I) and
CC   (II) may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate colon tumour associated protein (TCAP)
CC   expression, such as colonic cancer. For example, (I) and (II) may be
CC   used to treat disorders associated with decreased expression by
CC   rectifying mutations or deletions in a patient's genome that affect the
CC   activity of TCAPs by expressing inactive proteins or to supplement the
CC   patients own production of them. Additionally, (II) may be used to
CC   produce the TCAP proteins, by inserting the nucleic acids into a host
CC   cell culturing the cell to express the protein. (II) and its
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   polymerase chain reaction (PCR) and hybridisation assays to detect and
```

```
CC   quantitate the presence of similar nucleic acids in samples, and
CC   therefore which patients may be in need of restorative therapy. (I) may
CC   also be used as antigens in the production of antibodies against TCAPs
CC   and in assays to identify modulators of TCAP expression and activity.
CC   Anti-(I) antibodies and antagonists may also be used to down regulate
CC   TCAP expression and activity. The anti-(I) antibodies may also be used
CC   as diagnostic agents for detecting the presence of TCAPs in samples
CC   (e.g. by enzyme linked immunosorbant assay (ELISA)). AAI28460 to AAI29512
CC   and AAM24494 to AAM24523 represent nucleotide and amino acid sequences
CC   given in the exemplification of the present invention.
XX
SQ   Sequence   832 AA;

  Query Match             100.0%;  Score 4321;  DB 22;  Length 832;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches  832;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
```

```
RESULT 3
AAY99405
ID   AAY99405 standard; Protein; 807 AA.
XX
AC   AAY99405;
XX
DT   08-AUG-2000  (first entry)
XX
DE   Human PRO1340 (UNQ695) amino acid sequence SEQ ID NO:229.
XX
KW   Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW   transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
XX
OS   Homo sapiens.
XX
PN   WO200012708-A2.
XX
PD   09-MAR-2000.
XX
PF   01-SEP-1999;   99WO-US20111.
XX
PR   01-SEP-1998;   98US-0098716.
PR   01-SEP-1998;   98US-0098749.
PR   01-SEP-1998;   98US-0098750.
PR   02-SEP-1998;   98US-0098803.
PR   02-SEP-1998;   98US-0098821.
PR   02-SEP-1998;   98US-0098843.
PR   09-SEP-1998;   98US-0099536.
PR   09-SEP-1998;   98US-0099596.
PR   09-SEP-1998;   98US-0099598.
PR   09-SEP-1998;   98US-0099602.
PR   09-SEP-1998;   98US-0099642.
PR   10-SEP-1998;   98US-0099741.
PR   10-SEP-1998;   98US-0099754.
PR   10-SEP-1998;   98US-0099763.
PR   10-SEP-1998;   98US-0099792.
PR   10-SEP-1998;   98US-0099808.
PR   10-SEP-1998;   98US-0099812.
PR   10-SEP-1998;   98US-0099815.
PR   10-SEP-1998;   98US-0099816.
PR   15-SEP-1998;   98US-0100385.
PR   15-SEP-1998;   98US-0100388.
PR   15-SEP-1998;   98US-0100390.
PR   16-SEP-1998;   98US-0100584.
PR   16-SEP-1998;   98US-0100627.
PR   16-SEP-1998;   98US-0100661.
PR   16-SEP-1998;   98US-0100662.
PR   16-SEP-1998;   98US-0100664.
PR   17-SEP-1998;   98US-0100683.
PR   17-SEP-1998;   98US-0100684.
PR   17-SEP-1998;   98US-0100710.
PR   17-SEP-1998;   98US-0100711.
PR   17-SEP-1998;   98US-0100919.
PR   17-SEP-1998;   98US-0100930.
PR   18-SEP-1998;   98US-0100848.
PR   18-SEP-1998;   98US-0100849.
PR   18-SEP-1998;   98US-0101014.
PR   18-SEP-1998;   98US-0101068.
PR   18-SEP-1998;   98US-0101071.
PR   22-SEP-1998;   98US-0101279.
PR   23-SEP-1998;   98US-0101471.
PR   23-SEP-1998;   98US-0101472.
PR   23-SEP-1998;   98US-0101474.
PR   23-SEP-1998;   98US-0101475.
PR   23-SEP-1998;   98US-0101476.
PR   23-SEP-1998;   98US-0101477.
PR   23-SEP-1998;   98US-0101479.
PR   24-SEP-1998;   98US-0101738.
PR   24-SEP-1998;   98US-0101741.
PR   24-SEP-1998;   98US-0101743.
PR   24-SEP-1998;   98US-0101915.
PR   24-SEP-1998;   98US-0101916.
```

PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102487.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102684.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 14-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 20-OCT-1998; 98US-0105002.
PR 21-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0108500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 10-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.
PR 17-NOV-1998; 98US-0108779.
PR 17-NOV-1998; 98US-0108787.
PR 17-NOV-1998; 98US-0108788.
PR 17-NOV-1998; 98US-0108801.
PR 17-NOV-1998; 98US-0108802.
PR 17-NOV-1998; 98US-0108806.
PR 17-NOV-1998; 98US-0108807.
PR 17-NOV-1998; 98US-0108867.
PR 17-NOV-1998; 98US-0108925.
PR 18-NOV-1998; 98US-0108848.
PR 18-NOV-1998; 98US-0108849.
PR 18-NOV-1998; 98US-0108850.
PR 18-NOV-1998; 98US-0108851.
PR 18-NOV-1998; 98US-0108852.
PR 18-NOV-1998; 98US-0108858.
PR 18-NOV-1998; 98US-0108904.
XX
PA (GETH) GENENTECH INC.

XX
PI Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;
XX
DR WPI; 2000-237871/20.
DR N-PSDB; AAA37087.
XX
PT New mammalian DNA sequences encoding transmembrane, receptor or
PT secreted PRO polypeptides, useful for screening of potential peptide or
PT small molecule inhibitors of the relevant receptor/ligand interactions
XX
PS Claim 12; Fig 132; 773pp; English.
XX
CC AAA37022 to AAA37144 encode the new isolated human transmembrane,
CC receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The
CC transmembrane and receptor PRO proteins can be used for screening of
CC potential peptide or small molecule inhibitors of the relevant
CC receptor/ligand interactions. The polypeptides and nucleotide sequences
CC encoding then have various industrial applications, including uses as
CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
CC PCR primers and hybridisation probes used in the isolation of the PRO
CC polypeptides from the present invention.
XX
SQ Sequence 807 AA;

Query Match 21.0%; Score 908.5; DB 21; Length 807;
Best Local Similarity 30.2%; Pred. No. 1.1e-64;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

```
Qy      62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:    |:  | :: : | |   |||||| :: : |||     :| :: || |: : |
Db      57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEQAEYQLQVTLEMQDGHVLWGPQPVLVHV 116

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : : :||| |||:: |:| |:| | |  | : |: | |   :
Db     117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRFHILSQAPAQPS 176

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
           ||:  : ||:|: :||  |: |   :| |:: ||||| |: :    | :|:: : |
Db     177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKDMGDQA-SGHQATATVEVSIIE 235

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           : | : :|: : ||    :|  : || |:     | |   |  |  || :: ||:|||:
Db     236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFEVNAEGNLYVTR 292

Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
           |||| :  |:    }:: :|:  : |||:|| | | ||| | ||       : |    |
Db     293 ELDREAQAEYLLQVRAQNSHGEDYAAPLELHVLVMDENDNVPICPPRDPTVSIPELSPPG 352

Qy     357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQLAKQSLKK-QD 414
           :  |:| | |   : || : |:::   |:  ::|  | :   :| : |    |:  |:
Db     353 TEVTRLSAEDADAPGSPNSHVVYQLLSPEPEDGVEGRAFQVDPTSGSVTLGVLPLRAGQN 412

Qy     415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTI 472
           | ::::  :  | : | |:: | |||  | |  |  | ::| ||   |: :  :
Db     413 ILLLVLAMDLAGAEGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472

Qy     473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIK--KPLDFETAAVSN 528
           | ||| ||   : ::: : | :||:||  |:| :|  ::|:| ::  | | :| |
Db     473 TAIDADLEP---AFRLMDFAIERGDTEGTFGLDWEP--DSGHVRLRLCKNLSYEAAPSHE 527

Qy     529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
           :|   ::   || |       : |  |::|  |   |:  |  ::| |       |: :  :
Db     528 VVVVVQSVAKLV-GPGPGPGATATVTVLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586

Qy     589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647
           ||    : :||  |: ||| |:  :||: :   |   : |  | | | | :
Db     587 QPSDPISRTLRFSLVNDSEGWLCIEKFSGEVHTAQSLQGAQPGDTYTVLVEAQDTA---- 642

Qy     648 SSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF-EATDDDQHLFRGPHFTFS 706
             | |  |                : | |    | ::   : | |     || ::|:
Db     643 -------LTLAPVPSQ-----------YLCTPRQDHGLIVSGPSKDPDLASGHGP-YSFT 683

Qy     707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
```

```
          || : ::| || :  :||:|| |:      | | ::: : ::   :      : : |  |
Db    684 LGPNPTVQRDWRLQTLNGSHAYLTLALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740

Qy    766 SC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQA 824
          | ||| | |   |:||   ||||: ||: ||| | ::|      : ||  : | :
Db    741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGTLVAIGIFLILIFTHWTMSRKKDPDQPADS 800

Qy    825 SEVK 828
          :|
Db    801 VPLK 804


RESULT 4
AAB87574
ID   AAB87574 standard; Protein; 807 AA.
XX
AC   AAB87574;
XX
DT   15-MAY-2001  (first entry)
XX
DE   Human PRO1340.
XX
KW   Human; PRO protein; mapping.
XX
OS   Homo sapiens.
XX
PN   WO200116318-A2.
XX
PD   08-MAR-2001.
XX
PF   24-AUG-2000; 2000WO-US23328.
XX
PR   01-SEP-1999;   99WO-US20111.
PR   15-SEP-1999;   99WO-US21090.
PR   07-DEC-1999;   99US-0169495.
PR   09-DEC-1999;   99US-0170262.
PR   11-JAN-2000; 2000US-0175481.
PR   18-FEB-2000; 2000WO-US04341.
PR   18-FEB-2000; 2000WO-US04342.
PR   22-FEB-2000; 2000WO-US04414.
PR   01-MAR-2000; 2000WO-US05601.
PR   03-MAR-2000; 2000US-0187202.
PR   25-APR-2000; 2000US-0199397.
PR   22-MAY-2000; 2000WO-US14042.
PR   05-JUN-2000; 2000US-0209832.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Eaton DL,  Filvaroff E,  Gerritsen ME,  Goddard A,  Godowski PJ;
PI   Grimaldi CJ,  Gurney AL,  Watanabe CK,  Wood WI;
XX
DR   WPI; 2001-183260/18.
DR   N-PSDB; AAF92106.
XX
PT   Eighty four nucleic acids encoding PRO polypeptides, useful in
PT   molecular biology, including use as hybridization probes, and in
PT   chromosome and gene mapping. -
XX
PS   Claim 12; Fig 98; 278pp; English.
XX
CC   The present sequence is a human PRO polypeptide (secreted and
CC   transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
CC   anti-PRO antibodies are useful for preparation of a medicament useful in
CC   the treatment of a condition which is responsive to the PRO protein,
CC   agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
CC   employed as molecular weight markers for protein electrophoresis. The PRO
CC   coding sequence has applications in molecular biology, including use as
CC   hybridisation probes, and in chromosome and gene mapping.
XX
SQ   Sequence   807 AA;

  Query Match             21.0%;  Score 908.5;  DB 22;  Length 807;
  Best Local Similarity   30.2%;  Pred. No. 1.1e-64;

  Matches  237;  Conservative  159;  Mismatches  335;  Indels   53;  Gaps   21;

Qy     62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
          |:|:    |:  | :: : | |   |||||| :: : |||     :| :: || |: : |
Db     57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEQAEYQLQVTLEMQDGHVLWGPQPVLVHV 116

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
          || ||  | | |: |   : : :||| |||:: |:| |:| | |  | : |: | |   :
Db    117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRFHILSQAPAQPS 176

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
             ||:  : ||::|: :||  |: |   :| |:: ||||| |: :    | :|:: : |
Db    177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKDMGDQA-SGHQATATVEVSIIE 235

Qy    237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
          : | : :|: : ||    :|  : || |:     | |   |  | |  || :: ||::|||:
Db    236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFEVNAEGNLYVTR 292

Qy    297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
          |||| :  |:    |:: :|:  : |||:|| | | ||| | ||       : |    |
Db    293 ELDREAQAEYLLQVRAQNSHGEDYAAPLELHVLVMDENDNVPICPPRDPTVSIPELSPPG 352

Qy    357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQLAKQSLKK-QD 414
            :  |:| | |   : || : |:::   |:  ::|  | :   :| : |    |:  |:
Db    353 TEVTRLSAEDADAPGSPNSHVVYQLLSPEPEDGVEGRAFQVDPTSGSVTLGVLPLRAGQN 412

Qy    415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTI 472
               | ::::  :  | : | |:: | ||||  | |  |  | ::| ||   |: :  :
Db    413 ILLLVLAMDLAGAEGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472

Qy    473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIK--KPLDFETAAVSN 528
          | ||| ||   : ::: : | :||:||  |:| :|  ::|:| ::  | | :| |
Db    473 TAIDADLEP---AFRLMDFAIERGDTEGTFGLDWEP--DSGHVRLRLCKNLSYEAAPSHE 527

Qy    529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
          :|   ::   || |       : |  |::|  |   |:  |  ::| |      |: :  :
Db    528 VVVVVQSVAKLV-GPGPGPGATATVTVLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586

Qy    589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647
             ||    : :||  |: ||| |:  :||: :   |   : |  | | | | :
Db    587 QPSDPISRTLRFSLVNDSEGWLCIEKFSGEVHTAQSLQGAQPGDTYTVLVEAQDTA---- 642

Qy    648 SSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF-EATDDDQHLFRGPHFTFS 706
                | |  |                : | |    | ::   : | |     || ::|:
Db    643 -------LTLAPVPSQ-----------YLCTPRQDHGLIVSGPSKDPDLASGHGP-YSFT 683

Qy    707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
          || : ::| || :  :||:|| |:      | | ::: : ::   :      : : |  |
Db    684 LGPNPTVQRDWRLQTLNGSHAYLTLALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740

Qy    766 SC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQA 824
          | ||| | |   |:||   |||||: ||: ||| | ::|      : ||  : | :
Db    741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGTLVAIGIFLILIFTHWTMSRKKDPDQPADS 800

Qy    825 SEVK 828
          :|
Db    801 VPLK 804


RESULT 5
AAB66154
ID   AAB66154 standard; protein; 807 AA.
XX
AC   AAB66154;
XX
DT   02-APR-2001  (first entry)
XX
DE   Protein of the invention #66.
XX
KW   Secreted; transmembrane; gene therapy.
XX
OS   Unidentified.
XX
```

PN WO200078961-A1.
XX
PD 28-DEC-2000.
XX
PF 18-FEB-2000; 2000WO-US04342.
XX
PR 23-JUN-1999; 99US-0141037.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 02-DEC-1999; 99WO-US28551.
PR 16-DEC-1999; 99WO-US30095.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
PI Watanabe CK, Williams PM, Wood WI;
XX
DR WPI; 2001-071395/08.
XX
PT Secreted and transmembrane proteins and nucleic acids designated PRO,
PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -
XX
PS Claim 1; Fig 132; 787pp; English.
XX
CC The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.
XX
SQ Sequence 807 AA;

Query Match 21.0%; Score 908.5; DB 22; Length 807;
Best Local Similarity 30.2%; Pred. No. 1.1e-64;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

```
Qy     62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
Db     57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEQAEYQLQVTLEMQDGHVLWGPQPVLVHV 116

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
Db    117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRFHILSQAPAQPS 176

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
Db    177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKDMGDQA-SGHQATATVEVSIIE 235

Qy    237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
Db    236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFEVNAEGNLYVTR 292

Qy    297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
Db    293 ELDREAQAEYLLQVRAQNSHGEDYAAPLELHVLVMDENDNVPICPPRDPTVSIPELSPPG 352

Qy    357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQLAKQSLKK-QD 414
Db    353 TEVTRLSAEDADAPGSPNSHVVYQLLSPEPEDGVEGRAFQVDPTSGSVTLGVLPLRAGQN 412

Qy    415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTI 472
Db    413 ILLLVLAMDLAGAEGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472

Qy    473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIK--KPLDFETAAVSN 528
Db    473 TAIDADLEP---AFRLMDFAIERGDTEGTFGLDWEP--DSGHVRLRLCKNLSYEAAPSHE 527

Qy    529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
Db    528 VVVVVQSVAKLV-GPGPGPGATATVTVLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586

Qy    589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647
Db    587 QPSDPISRTLRFSLVNDSEGWLCIEKFSGEVHTAQSLQGAQPGDTYTVLVEAQDTA---- 642

Qy    648 SSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF-EATDDDQHLFRGPHFTFS 706
Db    643 -------LTLAPVPSQ-----------YLCTPRQDHGLIVSGPSKDPDLASGHGP-YSFT 683

Qy    707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
Db    684 LGPNPTVQRDWRLQTLNGSHAYLTLALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740

Qy    766 SC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQA 824
Db    741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGTLVAIGIFLILIFTHWTMSRKKDPDQPADS 800

Qy    825 SEVK 828
Db    801 VPLK 804
```

RESULT 6
AAU83685
ID AAU83685 standard; Protein; 807 AA.
XX
AC AAU83685;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human PRO protein, Seq ID No 188.
XX
KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
KW breast cancer; prostate tumour; rectal tumour; liver tumour;
KW pericyte cell proliferation; chondrocyte cell proliferation;
KW tumour necrosis factor-alpha.
XX
OS Homo sapiens.
XX
PN WO200208288-A2.
XX
PD 31-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-US21066.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220585P.
PR 25-JUL-2000; 2000US-220605P.
PR 25-JUL-2000; 2000US-220607P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220638P.
PR 25-JUL-2000; 2000US-220664P.
PR 25-JUL-2000; 2000US-220666P.
PR 26-JUL-2000; 2000US-220893P.
PR 28-JUL-2000; 2000WO-US20710.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 15-SEP-2000; 2000US-000000P.
PR 10-NOV-2000; 2000WO-US30873.
PR 28-NOV-2000; 2000US-253646P.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 10-MAY-2001; 2001US-0854280.

PR 25-MAY-2001; 2001WO-US17092.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WI;
XX
DR WPI; 2002-172001/22.
DR N-PSDB; ABK33629.
XX
PT One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT useful for treating a PRO related disorder and for diagnosing tumours
PT such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
PT tumour or liver tumour -
XX
PS Claim 11; Figure 188; 359pp; English.
XX
CC The invention relates to one hundred and twenty two nucleic acids
CC encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
CC encode human secreted proteins. The PRO nucleic acids, polypeptides,
CC agonists and antagonists are useful for treating a PRO related disorder.
CC The PRO polypeptides are useful for diagnosing tumours, especially lung
CC cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
CC liver tumour. The PRO polypeptides are useful for stimulating the
CC proliferation of, or gene expression, in pericyte cells, for stimulating
CC the proliferation or differentiation of chondrocyte cells, for
CC stimulating the release of tumour necrosis factor-alpha from human blood,
CC for stimulating or inhibiting the proliferation of normal human dermal
CC fibroblast cells. The PRO polypeptide may also be used as molecular
CC weight markers and for tissue typing. The PRO nucleic acids have
CC applications in molecular biology, including use as hybridisation probes,
CC and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO
CC protein sequences of the invention.
XX
SQ Sequence 807 AA;

Query Match 21.0%; Score 908.5; DB 23; Length 807;
Best Local Similarity 30.2%; Pred. No. 1.1e-64;
Matches 237; Conservative 159; Mismatches 335; Indels 53; Gaps 21;

```
Qy      62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:    |:  | :: : | |   ||||||  :: : |||     :| :: || |: : |
Db      57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEQAEYQLQVTLEMQDGHVLWGPQPVLVHV 116

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : : :||| |||:: |:| |:| | |  | : |: | |   :
Db     117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRFHILSQAPAQPS 176

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
              ||:  : ||::|: :||  |: |   :| |:: ||||| |: :    | :|:: : |
Db     177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKDMGDQA-SGHQATATVEVSIIE 235

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           : | : :|: : ||    :|  : || |:     | |   |  |  || :: ||::|||:
Db     236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFEVNAEGNLYVTR 292

Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
           |||| :  |:    |:: :|:  : |||:|| | | ||| | ||       : |    |
Db     293 ELDREAQAEYLLQVRAQNSHGEDYAAPLELHVLVMDENDNVPICPPRDPTVSIPELSPPG 352

Qy     357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQLAKQSLKK-QD 414
             :  |:| | |   : || : |:::   |:  ::|  | :   :| : |    |:  |:
Db     353 TEVTRLSAEDADAPGSPNSHVVYQLLSPEPEDGVEGRAFQVDPTSGSVTLGVLPLRAGQN 412

Qy     415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTI 472
           | ::::  :  | : | |:: | ||||  | |  |  | ::| ||   |: :  :
Db     413 ILLLVLAMDLAGAEGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472

Qy     473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIK--KPLDFETAAVSN 528
           | ||| ||   : ::: : | :||:||  |:| :|  ::|:| ::  | | :| |
Db     473 TAIDADLEP---AFRLMDFAIERGDTEGTFGLDWEP--DSGHVRLRLCKNLSYEAAPSHE 527

Qy     529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
           :|   ::   || |      : |  |::|  |   |:  |  ::| |      |: :  :
Db     528 VVVVVQSVAKLV-GPGPGPGATATVTVLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586

Qy     589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647
             ||    : :||  |: ||| |:  :||: :   |   : |  | | | | :
Db     587 QPSDPISRTLRFSLVNDSEGWLCIEKFSGEVHTAQSLQGAQPGDTYTVLVEAQDTA---- 642

Qy     648 SSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF-EATDDDQHLFRGPHFTFS 706
                 | |  |              : | |    | ::   : | |     || ::|:
Db     643 -------LTLAPVPSQ----------YLCTPRQDHGLIVSGPSKDPDLASGHGP-YSFT 683

Qy     707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
           || : ::| || :  :||:|| |:      | | ::: : ::   :      : : |  |
Db     684 LGPNPTVQRDWRLQTLNGSHAYLTLALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740

Qy     766 SC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQA 824
           | ||| | |  |   |:||  |||||: ||: ||| | ::|      : ||  : | :
Db     741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGTLVAIGIFLILIFTHWTMSRKKDPDQPADS 800

Qy     825 SEVK 828
             :|
Db     801 VPLK 804
```

RESULT 7
AAU09959
ID AAU09959 standard; Protein; 830 AA.
XX
AC AAU09959;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human cadherin-2 (CDHN-2) protein.
XX
KW Cadherin; CDHN-2; nootropic; neuroprotective; cardiant; human;
KW antiinflammatory; gastric; immunostimulant; cytostatic; immunogen;
KW Alzheimer's disease; Parkinson's disease; multiple sclerosis;
KW atherosclerosis; angiogenesis; musculoskeletal; ataxia; myotonia;
KW gastrointestinal; gastritis; insulin dependent diabetes mellitus;
KW Crohn's disease; inflammatory; asthma; rheumatoid arthritis; lupus;
KW cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note= "signal peptide"
FT Protein 22..830
FT /note= "Mature CDHN-2"
FT Domain 540..557
FT /note= "Transmembrane domain"
FT Domain 571..588
FT /note= "Transmembrane domain"
FT Domain 789..813
FT /note= "Transmembrane domain"
XX
PN WO200179293-A2.
XX
PD 25-OCT-2001.
XX
PF 18-APR-2001; 2001WO-US12687.
XX
PR 18-APR-2000; 2000US-198466P.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Curtis RAJ;
XX
DR WPI; 2001-611722/70.
DR N-PSDB; AAS15602.
XX
PT Isolated cadherin polypeptide useful for the treatment of central
PT nervous system, cardiovascular, musculoskeletal, gastrointestinal,

```
PT   inflammatory, immune system and cell proliferation disorders -
XX
PS   Claim 13; Figure 8; 147pp; English.
XX
CC   This sequence represents the cadherin-2 (CDNH-2) protein of the
CC   invention. This invention comprises the nucleotide and protein
CC   sequences of two members of the cadherin family of membrane
CC   glycoproteins CDHN-1 and CDHN-2. Also included in the invention is
CC   a method for producing the CDHN proteins by culturing a transfected
CC   host cell line with a vector containing the CDHN proteins and an
CC   antibody which selectively binds to the protein. The CDHN proteins
CC   may have nootropic, neuroprotective, cardiant, antiinflammatory,
CC   gastric, immunostimulant and cytostatic activity and can be used
CC   as cadherin-modulators using antisense-therapy or gene-therapy.
CC   These proteins are useful for identifying compounds which bind
CC   or modulate CDHN-1. The proteins, nucleic acids and antibodies
CC   are useful for treating a subject with a disorder characterised
CC   by aberrant or unwanted cadherin protein or nucleic acid activity.
CC   These disorders include central nervous system (e.g. Alzheimer's
CC   disease, Parkinson's disease, multiple sclerosis), cardiovascular,
CC   (e.g. atherosclerosis, angiogenesis), musculoskeletal (e.g. ataxia,
CC   myotonia) gastrointestinal (e.g. gastritis, insulin dependent
CC   diabetes mellitus, Crohn's disease) (e.g. inflammatory and
CC   immune system (e.g. asthma, rheumatoid arthritis, lupus) or cell
CC   proliferation disorders (e.g. cancers, leukaemia) and many other
CC   diseases/disorders listed in the specification.
XX
SQ   Sequence   830 AA;

  Query Match             20.3%;  Score 877.5;  DB 22;  Length 830;
  Best Local Similarity   30.5%;  Pred. No. 3.7e-62;
  Matches  245;  Conservative  141;  Mismatches  353;  Indels   63;  Gaps   22;

Qy      62 LTGETD----NIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:::    | | :: : | |   | |||| :: : |||     :| |: ||  :|: |
Db      59 LSGDSNTADQNTFAVDTDSGFLVATRTLDREEKAEYQLQVTLESEDGRILWGPQLVTVHV 118

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : | :||| |||:: |:| | | | |  | : |: | |
Db     119 KDENDQVPQFSQAIYRAQLSQGTRPGVPFLFLEASDGDAPGTANSDLRFHILSQSPPQPL 178

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
              ||::   ||::|:  ||  |: |   :| |:: ||||| |     :  | |:| : |
Db     179 PDMFQLDPHLGALALSPSGSTSLDHALEETYQLLVQVKDMGDQPSGHQAIAT-VEISIVE 237

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           | |   :|| : ||    :|  | || |:     | |   |  |  || :| || ::||
Db     238 NSWAPLEPVHLAENLKVVYPHSIAQVHWSGGDVHYQL---ESQPPGPFDVDTEGMLHVTM 294

Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
           ||||  :  |     |:: :|:  : |||: | | | ||| | |        : |    |
Db     295 ELDREAQAEYQLQVRAQNSHGEDYAEPLELQVVVMDENDNAPVCSPHDPTVNIPELSPPG 354

Qy     357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQL------AKQS 409
             |  |:| | |   : || : |:::   |:   :   | :   :| : |      | ||
Db     355 TEIARLSAEDLDAPGSPNSHIVYQLLSPEPEEGAENKAFELDPTSGSVTLGTAPLHAGQS 414

Qy     410 LKKQDTPQYNLTIEV--SDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGS 467
           :  |      | :::  |:    : | | : | |:|:  | |  |  | :|| ||   |:
Db     415 ILLQ-----VLAVDLAGSESGLSSTCEVTVMVTDVNNHAPEFINSQIGPVTLPEDVKPGA 469

Qy     468 TILTIQATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAA 525
           : |: |||| ||   : ::: : | :|| ||   :  :| ::   : ::| | :| |
Db     470 LVATLMATDADLEP---AFRLMDFAIEEGDPEGIFDLSWEPDSDHVQLRLRKNLSYEAAP 526

Qy     526 VSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKV 585
             :|    | | || |     :: |  |::|  |    :  |  ::  :      |: :
Db     527 DHKVVVVVSNIEELV-GPGPGPAATATVTILVERVVAPLKLDQESYETSIPVSTPAGSLL 585

Qy     586 GNVTAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVA--TEV 642
             :   ||    : :||  |: ||| |  |:||: :   |   : |  | | | |  |:
Db     586 LTIQPSDPMSRTLRFSLVNDSEGWLCIKEVSGEVHTAQSLQGAQPGDTYTVLVEAQDTDK 645

Qy     643 GGSSLSSVSEFHLI------LMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQH 696
           | | |:    | :       : ::  | |          | |    | ::   ::|
Db     646 PGLSTSATVVIHFLKASPVPALTLSAGPSR--------HLCTPRQDYGVVVSGVSEDPDL 697

Qy     697 LFRGPHFTFSLG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRI-NDGGRPPL 754
           |   ::|:|| : ::| || :  :| :|| |:      |   |:| : : :|     |
Db     698 ANRNGPYSFALGPNPTVQRDWRLQPLNDSHAYLTLALHWVEPGEYMVPVVVHHDTHMWQL 757

Qy     755 EGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRI--- 810
           :    : |  | | ||| | |  |   |:||   |||:|| ||  || || :||  :
Db     758 Q----VKVIVCRCNVEGQCMRKVGRMKGMPTKLSAVGVLLGTLAAIGFILILVFTHLALA 813

Qy     811 KKDKGKDNVESAQASEVKPLRS 832
           :||         | ::  ||::
Db     814 RKDLD-------QPADSVPLKA 828


RESULT 8
AAY70741
ID   AAY70741 standard; protein; 906 AA.
XX
AC   AAY70741;
XX
DT   24-JUL-2000  (first entry)
XX
DE   Human N-cadherin.
XX
KW   Human N-cadherin; Wnt antagonist; contraceptive; contraceptive vaccine;
KW   oocyte development; female primate contraception; oocyte viability;
KW   monoclonal antibody; Wnt signalling.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          160..724
FT                   /label= Extracellular_domain
FT   Domain          747..906
FT                   /label= Cytoplasmic_domain
XX
PN   WO200021555-A1.
XX
PD   20-APR-2000.
XX
PF   13-OCT-1999;   99WO-US23640.
XX
PR   15-OCT-1998;   98US-0104355.
XX
PA   (HARD ) HARVARD COLLEGE.
XX
PI   Mcmahon AP,  Parr BA,  Vaino S;
XX
DR   WPI; 2000-317845/27.
XX
PT   Contraceptive composition for inhibiting oocyte development in a female
PT   primate comprises a Wnt polypeptide antagonist  -
XX
PS   Example 3; Page 27-28; 57pp; English.
XX
CC   The patent discloses a method of female primate contraception comprising
CC   administering an antagonist of a Wnt polypeptide, inhibiting oocyte
CC   development. Wnt polypeptides are useful for promotive maturation of an
CC   immature oocyte. Wnt polypeptides are also useful for increasing the
CC   number of mature oocytes and to enhance oocyte viability. Soluble
CC   fragments of Wnt polypeptides have the ability to inhibit Wnt signalling,
CC   e.g., by blocking binding of a naturally-occurring Wnt protein to its
CC   receptor. They may be used to generate monoclonal antibodies which can
CC   inhibit oocyte development. The present sequence is the human N-cadherin.
CC   Signal transduction by beta-catenin is mediated by binding to the
CC   cytoplasmic domain of a cadherin. Dominant negative mutants of N-cadherin
CC   which inhibit Wnt-1 class signals can also be used as Wnt antagonists.
XX
SQ   Sequence   906 AA;
```

```
  Query Match             16.2%;  Score 699.5;  DB 21;  Length 906;
  Best Local Similarity   27.4%;  Pred. No. 1.3e-47;
  Matches  210;  Conservative  122;  Mismatches  331;  Indels  103;  Gaps   21;

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           ||::: :|  |  |:       |    :   | ::   |      :| :|   |   |: :
Db      47 KDVHEGQP-LLNVKF-----SNCNGKRKVQYESSEPADFKVDEDGMVY--AVRSFPLSSE 98

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
             | |            :|:    :     : :|:|     :| |  ::  |: ||
Db      99 HAKFLIY-----------AQDKETQEKWQVAVKLSLKPT--LTEESVKESAEVEEIVFP 144

Qy     235 -----------TENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQ 270
                       :  |  | |: : |||  | | :: ::| :             |||
Db     145 RQFSKHSGHLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA- 202

Qy     271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
             :: |   | |:   | : ||:|||||:   :   | | |  |  :  |::| :
Db     203 ------DQPPTGIFIINPISGQLSVTKPLDREQIARFHLRAHAVDINGNQVENPIDIVIN 256

Qy     330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
           | |:||| |     |    | |  : |  : |:|| | |: |  |  | |||| | |  |
Db     257 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316

Qy     390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
             :| |    | :      | ::   || | |: :| :             | | |:|
Db     317 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVN 376

Qy     444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502
           |  | |   : || :   :  :  :  :  || |:| | :   :| |  ||  ||  :
Db     377 DNPPEFTAMTFYGEV---PENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAI 433

Qy     503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
           |||::| | | : ||:|||  :  :   |||  ||  |:::   | |  :: | ||||
Db     434 QTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 493

Qy     563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
             | |: :    :  | :  || :   ||:||:     :| |:   |   |||| | |:|
Db     494 NPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQI 553

Qy     620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675
           ::| ||||:      : |    :|:: |   :|      : |:|:||| |::
Db     554 TTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL------- 606

Qy     676 FCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727
             |  |       | |:   | | |     ||  |   |   ::: :| ::::||  |:
Db     607 ---PQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQ 663

Qy     728 LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV- 785
           |: :    |   | | | | | | ||   | | |  | |    | :  :
Db     664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIV 716

Qy     786 --GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
             |:  | ::  || | |:| :| : :   | :|    |:   : |
Db     717 GAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762
```

```
RESULT 9
ABG14316
ID   ABG14316 standard; Protein; 906 AA.
XX
AC   ABG14316;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #14307.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS78503.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 44675; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   906 AA;
```

```
  Query Match             16.2%;  Score 699.5;  DB 22;  Length 906;
  Best Local Similarity   27.4%;  Pred. No. 1.3e-47;
  Matches  210;  Conservative  122;  Mismatches  331;  Indels  103;  Gaps   21;

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           ||::: :|  |  |:       |    :   | ::   |      :| :|   |   |: :
Db      47 KDVHEGQP-LLNVKF-----SNCNGKRKVQYESSEPADFKVDEDGMVY--AVRSFPLSSE 98

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
             | |            :|:    :     : :|:|     :| |  ::  |: ||
Db      99 HAKFLIY-----------AQDKETQEKWQVAVKLSLKPT--LTEESVKESAEVEEIVFP 144

Qy     235 -----------TENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQ 270
                       :  |  | |: : |||  | | :: ::| :             |||
Db     145 RQFSKHSGHLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA- 202

Qy     271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
             :: |   | |:   | : ||:|||||:   :   | | |  |  :  |::| :
Db     203 ------DQPPTGIFIINPISGQLSVTKPLDREQIARFHLRAHAVDINGNQVENPIDIVIN 256

Qy     330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
           | |:||| |     |    | |  : |  : |:|| | |: |  |  | |||| | |  |
Db     257 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316

Qy     390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
             :| |    | :      | ::   || | |: :| :             | | |:|
Db     317 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVN 376
```

```
Qy    444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502
Db    377 DNPPEFTAMTFYGEV---PENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAI 433

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
Db    434 QTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 493

Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
Db    494 NPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQI 553

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675
Db    554 TTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL------- 606

Qy    676 FCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727
Db    607 ---PQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQ 663

Qy    728 LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV- 785
Db    664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIV 716

Qy    786 --GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
Db    717 GAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762


RESULT 10
ABB81474
ID   ABB81474 standard; Protein; 906 AA.
XX
AC   ABB81474;
XX
DT   30-AUG-2002  (first entry)
XX
DE   Human N-cadherin protein SEQ ID NO:46.
XX
KW   Cadherin; alpha-catenin; cancer; beta-catenin binding domain; melanoma;
KW   o-catenin; colon cancer.
XX
OS   Homo sapiens.
XX
PN   US2002045591-A1.
XX
PD   18-APR-2002.
XX
PF   17-JUL-2001; 2001US-0905983.
XX
PR   26-MAY-1998;   98IL-0124650.
PR   26-MAY-1999;   99US-0318633.
XX
PA   (GEIG/) GEIGER B.
PA   (BENZ/) BEN-ZE'EV A.
PA   (SADO/) SADOT E.
XX
PI   Geiger B,  Ben-Ze'ev A,  Sadot E;
XX
DR   WPI; 2002-499105/53.
DR   N-PSDB; ABN89391.
XX
PT   New construct encoding soluble cytoplasmic portion of cadherin
PT   including beta catenin binding domain useful in treating cancer
PT   associated with high beta-catenin activity e.g. colon cancer and
PT   melanoma  -
XX
PS   Example 3; Page 45-48; 102pp; English.
XX
CC   The present invention describes a pharmaceutical composition for treating
CC   cancer associated with abnormally high beta-catenin activity. The
CC   pharmaceutical composition comprises a gene therapy vehicle harbouring a
CC   polynucleotide that contains: (i) a nucleotide sequence encoding a
CC   soluble cytoplasmic portion of a cadherin which lacks a transmembrane
CC   portion and an extracellular portion of the cadherin, and includes a
CC   beta-catenin binding domain; and (b) an upstream promoter for directing
CC   expression of the soluble cytoplasmic portion of the cadherin in a
CC   mammalian cell. Also described is a pharmaceutical composition for
CC   treating cancer associated with abnormally high activity levels of
CC   beta-catenin comprising a gene therapy vehicle harbouring a
CC   polynucleotide that contains: (a) a nucleotide sequence encoding an
CC   o-catenin; and (b) an upstream promoter for directing expression of the
CC   o-catenin in a mammalian cell. The pharmaceutical compositions have
CC   cytostatic activity and can be used in the suppression of
CC   beta-catenin-mediated transactivation. They can be used for treating
CC   cancers associated with abnormally high activity levels of beta-catenin
CC   such as colon cancers and melanomas, by reducing these high activity
CC   levels of beta-catenin in mammalian cells. The present sequence
CC   represents human N-cadherin which is used in the exemplification of the
CC   present invention.
XX
SQ   Sequence   906 AA;

  Query Match             16.2%;  Score 699.5;  DB 23;  Length 906;
  Best Local Similarity   27.4%;  Pred. No. 1.3e-47;
  Matches  210;  Conservative  122;  Mismatches  331;  Indels  103;  Gaps   21;

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
Db     47 KDVHEGQP-LLNVKF-----SNCNGKRKVQYESSEPADFKVDEDGMVY--AVRSFPLSSE 98

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
Db     99 HAKFLIY------------AQDKETQEKWQVAVKLSLKPT--LTEESVKESAEVEEIVFP 144

Qy    235 ------------TENIWKAPKPVEMVENSTDPHPIKITQVRWN------------DPGAQ 270
Db    145 RQFSKHSGHLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYTVTGPGA- 202

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
Db    203 ------DQPPTGIFIINPISGQLSVTKPLDREQIARFHLRAHAVDINGNQVENPIDIVIN 256

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
Db    257 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
Db    317 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGIPTYGLSNTATAVITVTDVN 376

Qy    444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502
Db    377 DNPPEFTAMTFYGEV---PENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAI 433

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
Db    434 QTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 493

Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
Db    494 NPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQI 553

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675
Db    554 TTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL------- 606

Qy    676 FCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727
Db    607 ---PQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQ 663

Qy    728 LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV- 785
Db    664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIV 716
```

```
Qy    786 --GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
          |:  | ::  || | |:| :| : :   | :|    |:   : |
Db    717 GAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762
```

RESULT 11
ABB81471
ID ABB81471 standard; Protein; 912 AA.
XX
AC ABB81471;
XX
DT 30-AUG-2002 (first entry)
XX
DE Chicken N-cadherin protein SEQ ID NO:2.
XX
KW Cadherin; alpha-catenin; cancer; beta-catenin binding domain; melanoma;
KW o-catenin; colon cancer.
XX
OS Gallus gallus.
XX
PN US2002045591-A1.
XX
PD 18-APR-2002.
XX
PF 17-JUL-2001; 2001US-0905983.
XX
PR 26-MAY-1998; 98IL-0124650.
PR 26-MAY-1999; 99US-0318633.
XX
PA (GEIG/) GEIGER B.
PA (BENZ/) BEN-ZE'EV A.
PA (SADO/) SADOT E.
XX
PI Geiger B, Ben-Ze'ev A, Sadot E;
XX
DR WPI; 2002-499105/53.
DR N-PSDB; ABN89350, ABN89351.
XX
PT New construct encoding soluble cytoplasmic portion of cadherin
PT including beta catenin binding domain useful in treating cancer
PT associated with high beta-catenin activity e.g. colon cancer and
PT melanoma -
XX
PS Example 1; Page 19-21; 102pp; English.
XX
CC The present invention describes a pharmaceutical composition for treating
CC cancer associated with abnormally high beta-catenin activity. The
CC pharmaceutical composition comprises a gene therapy vehicle harbouring a
CC polynucleotide that contains: (i) a nucleotide sequence encoding a
CC soluble cytoplasmic portion of a cadherin which lacks a transmembrane
CC portion and an extracellular portion of the cadherin, and includes a
CC beta-catenin binding domain; and (b) an upstream promoter for directing
CC expression of the soluble cytoplasmic portion of the cadherin in a
CC mammalian cell. Also described is a pharmaceutical composition for
CC treating cancer associated with abnormally high activity levels of
CC beta-catenin comprising a gene therapy vehicle harbouring a
CC polynucleotide that contains: (a) a nucleotide sequence encoding an
CC o-catenin; and (b) an upstream promoter for directing expression of the
CC o-catenin in a mammalian cell. The pharmaceutical compositions have
CC cytostatic activity and can be used in the suppression of
CC beta-catenin-mediated transactivation. They can be used for treating
CC cancers associated with abnormally high activity levels of beta-catenin
CC such as colon cancers and melanomas, by reducing these high activity
CC levels of beta-catenin in mammalian cells. The present sequence
CC represents chicken N-cadherin which is used in the exemplification of
CC the present invention.
XX
SQ Sequence 912 AA;

Query Match 16.1%; Score 696; DB 23; Length 912;
Best Local Similarity 26.9%; Pred. No. 2.6e-47;
Matches 209; Conservative 117; Mismatches 328; Indels 124; Gaps 22;

```
Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183
          | :  :   | ::   |:| | |     |:                    |  :||  :
Db     41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy    184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENFSDTT 228
               : | :   |       | :      |:| :|   |           | :|:  :
Db     82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy    229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269
            |    |:|                  |  | |: : |||  | | :: ::|
Db    137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy    270 QYSLVDKEKLPRFPFSI-----DQ----------EGDIYVTQPLDREEKDAYVFYAVAK 313
               |::|     :|:     ||           | : ||:|||||:  ::   | |
Db    190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy    314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373
          |  |  :  |::| : | |:||| |     |    | |  : |  : |:|| | |: |
Db    246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy    374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428
          |  | |||: | |  |   :| |    | :      | ::   || | |: :| :
Db    306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy    429 -FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
                | | |:||  | |   : || :   :  :   :  :  || |:| | :
Db    366 GLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWN 422

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
          | :  ||  |:  : |||::| | | : ||:|||| :  :   ||| || |:::
Db    423 ARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISYSL 602
          | |  :: | ||||:| |  :    :  | :  |: :   ||:||    :   : ||
Db    483 PQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRYSK 542

Qy    603 RGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILM 658
            |   ||||| | |:| : | ||||:     : |    :|:: |   :|      : |:
Db    543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIYLL 602

Qy    659 DVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQND 715
          |:||| |::  |: |     |  |  | ::   | | |     || | | |     |:: :
Db    603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657

Qy    716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFR 774
          | : :|:| ||:|| |    |   | | | | | | |       | |  | | : | |
Db    658 WTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC-- 715

Qy    775 PAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
               | :  :   |:  | ::  || | |:| :| : :   | :|    |:   : |
Db    716 -----TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768

RESULT 12
ABB57233
ID  ABB57233 standard; Protein; 906 AA.
XX
AC  ABB57233;
XX
DT  07-MAR-2002 (first entry)
XX
DE  Mouse ischaemic condition related protein sequence SEQ ID NO:606.
XX
KW  Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
KW  vasospastic ischaemia; ischaemic condition; ischaemic disease.
XX
OS  Mus musculus.
XX
PN  WO200188188-A2.
XX
PD  22-NOV-2001.
XX
PF  18-MAY-2001; 2001WO-JP04192.
```

```
XX
PR   18-MAY-2000; 2000JP-0145977.
XX
PA   (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
XX
PI   Ishikawa K,  Asai S,  Takahashi Y,  Nagata T,  Ishii Y;
XX
DR   WPI; 2002-034733/04.
DR   N-PSDB; ABI99593.
XX
PT   Examining the ischemic condition (e.g. occlusive ischemia) by measuring
PT   expression levels of particular genes defined in the specification or
PT   by determining the expression profile of a gene group comprising these
PT   genes -
XX
PS   Claim 2; Page 1560-1564; 2690pp; English.
XX
CC   The present invention describes a method for examining ischaemic
CC   conditions, comprising measuring the expression levels of particular
CC   genes (I) in a test sample or determining the expression profile of a
CC   gene group in the sample comprising genes selected from (I). The method
CC   is useful for examining the ischaemic condition (e.g. compressive
CC   ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring
CC   expression levels of particular genes (ABI99202 to ABI99912, encoding
CC   the protein sequences in ABB57020 to ABB57374) or by determining the
CC   expression profile of a gene group comprising these genes. The
CC   expression levels or expression profiles produced by these genes are
CC   used as an indicator when screening for ischaemic condition-improving
CC   drugs or therapeutics for ischaemic diseases. ABI99913 and ABI99914
CC   represent PCR primers for a mouse ischaemic condition related sequence,
CC   which are used in the exemplification of the present invention.
XX
SQ   Sequence   906 AA;

  Query Match             16.1%;  Score 694.5;  DB 23;  Length 906;
  Best Local Similarity   27.6%;  Pred. No. 3.4e-47;
  Matches  211;  Conservative  122;  Mismatches  332;  Indels   99;  Gaps   23;

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           ||::: :|  |  |:    |:     :  | ::   |     :| :|   |   |:
Db      47 KDVHEGQP-LLNVKFSNCNRK-----RKVQYESSEPADFKVDEDGTVY--AVRSFPLTAE 98

Qy     177 VMYFQI---NNKTG-----AISLTREGSQELNPAKNP-SYNLVISVKDM---GGQSENSF 224
             | |   : :|      |::|:|| :    | | |     ::  : :    |  :
Db      99 QAKFLIYAQDKETQEKWQVAVNLSREPTLTEEPMKEPHEIEEIVFPRQLAKHSGALQRQK 158

Qy     225 SDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQYS 272
           |            |  | |: : |||  | | :: ::| :             |||
Db     159 RD-----------WVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA--- 202

Qy     273 LVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVK 331
           :: |   | |:   | : ||:|||||    :   | | |  |  :  |::| : |
Db     203 ----DQPPTGIFIINPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVINVI 258

Qy     332 DINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMD 391
           |:||| |     |   | |  : |  : |:|| | |: |  |  | |||: | |  |
Db     259 DMNDNRPEFLHQVWNGSVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTPSP 318

Qy     392 GLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQ 445
           :| |    | :      | ::   || | |: :| :              | | |:||
Db     319 NMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVNDN 378

Qy     446 IPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDT 504
           | |   : || :    :  :   :  : || |:| | :    | |  ||  ||  : |
Db     379 PPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWNAAYRISGGDPTGRFAILT 435

Qy     505 DPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAP 564
           ||::| | | : ||:|||  :  :   |||  ||  |:::   | |  :: | |||| |
Db     436 DPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNENP 495

Qy     565 QFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFS 621
           |: :    :  | :  || :  :||:||:     :| |:   |   ||||| | |:| :
Db     496 YFAPNPKIIRQEEGLHAGTMLTTLTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQITT 555
```

```
Qy     622 VAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFC 677
           :| ||||:     : |    :|:: |   :|       : |:|:||| |::
Db     556 IAVLDRESPYVQNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL--------- 606

Qy     678 HPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLS 729
           |  |        | |:   | | |      ||  |   |   ::: :| ::::||  |:|:
Db     607 -PQEAETCETPEPNSINIAALDYDIDPNAGPFAFDLPLSPVTIKRNWTINRLNGDFAQLN 665

Qy     730 TRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV--- 785
           :    |   | | | | | | ||   |  | |  | |   | |       | :  :
Db     666 LKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIVGA 718

Qy     786 GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
           |:  | ::  || | |:| :| : :   | :|    |:   : |
Db     719 GLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762


RESULT 13
AAW25658
ID   AAW25658 standard; Protein; 916 AA.
XX
AC   AAW25658;
XX
DT   04-NOV-1997  (first entry)
XX
DE   Human cadherin-4.
XX
KW   Human; cadherin; rat; calcium-dependent cell adhesion protein;
KW   superfamily; cytoskeleton; eatenin; cancer.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 78
FT                   /note= "Encoded by AAA"
XX
PN   US5646250-A.
XX
PD   08-JUL-1997.
XX
PF   17-APR-1992;   92US-0872643.
XX
PR   19-APR-1993;   93US-0049460.
PR   17-APR-1992;   92US-0872643.
PR   01-NOV-1994;   94US-0332638.
XX
PA   (DOHE-) DOHENY EYE INST.
XX
PI   Suzuki S;
XX
DR   WPI; 1997-362997/33.
DR   N-PSDB; AAT85433.
XX
PT   Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion
XX
PS   Example 2; Column 61-66; 56pp; English.
XX
CC   This sequence represents human cadherin-4.  The invention specifically
CC   provides details of human cadherin-5, -8, -11, -12 and -13, and rat
CC   cadherin-8, -5, -11 and -13.  Cadherins are calcium-dependent cell
CC   adhesion proteins.  They are glycosylated integral membrane proteins
CC   that have an N-terminal extracellular domain that determines binding
CC   specificity, a hydrophobic membrane spanning region and a C-terminal
CC   cytoplasmic domain, which is highly conserved among members of the
CC   superfamily.  The C-terminal domain interacts with the cytoskeleton
CC   through eatenins and other cytoskeleton-associated proteins.  The
CC   novel cadherin proteins may be used in the analysis of the role of
CC   cadherins in various cancers.  Sequence analysis of the cadherin
CC   proteins also allows investigation of the structure and function of
CC   cadherin.  The cadherin proteins may be isolated by using anti-cadherin
CC   antibodies.  These antibodies may also be used to modulate the activity
CC   of cadherin and to determine the tissue specific distribution of cadherin
```

```
CC   proteins.  Each subclass of cadherins has a unique tissue distribution
CC   pattern.
XX
SQ   Sequence   916 AA;

  Query Match             15.8%;  Score 684.5;  DB 18;  Length 916;
  Best Local Similarity   29.3%;  Pred. No. 2.3e-46;
  Matches  176;  Conservative  106;  Mismatches  288;  Indels   31;  Gaps   12;

Qy     239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
           |  | |: : |||  | | :: ::|    ||   :||:  |  :: |   |||:   | :
Db     171 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFSINSMSGRM 229

Qy     293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
           |||:|:|||| :|   | | |  |  :  |:::::| |:||| |  : |    | |
Db     230 YVTRPMDREEHASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNHPEFINQVYNCSVDEG 289

Qy     353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
            : |  : |:||:| |:  |||  : |||| |||: |   :| | :  | :         :
Db     290 SKPGTYVMTITANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349

Qy     413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
           :   || : :: :| :              | | |:||    |  | :       : ::
Db     350 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPSEFTASTFAG--EVPENSVE 407

Qy     467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
           : :  :   | |:| : :   :| || ||  |   | ||| || | | : | :|:|
Db     408 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 467

Qy     527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
             :     |  ||  |:: :  | |  |: : |:|||| |  :    :: | |  || :
Db     468 FMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPYFPSNHKLIRLEEGVPPGTVLT 527

Qy     587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
             :| ||:      : ||   |   || |:  |:| :|| ||||:     : |   :|
Db     528 TFSAVDPDRFMQQAVRYSKLSDPASWLHINATNGQITTVAVLDRESLYTKNNVYEATFLA 587

Qy     640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFR 699
           : |    |     : |:|:||| | |      :    |:|   :   | | | |
Db     588 ADNGIPPASGTGTLQIYLIDINDNAPELLPKEAQICERPNLNA---INITAADADVHPNI 644

Qy     700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIV 758
           ||: |       ::: :| ::::|| :|:|| |    |   | | | : | | |||
Db     645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPIIVTDSGNPPLSNTS 704

Qy     759 SLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--FIRIKKDKGK 816
           : |  | | :       |    :   |:  | ::  |: | |:| :|  |:   | : |
Db     705 IIKVKVCPCDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTMVLLFVMWMKRREK 761

Qy     817 D 817
           :
Db     762 E 762


RESULT 14
AAW13129
ID   AAW13129 standard; Protein; 916 AA.
XX
AC   AAW13129;
XX
DT   13-MAY-1997  (first entry)
XX
DE   Full length human cadherin-4.
XX
KW   Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;
KW   brain; human; antibody; purification; determination;
KW   tissue expression; binding antagonist; calcium ion.
XX
OS   Homo sapiens.
XX
PN   US5597725-A.
XX
PD   28-JAN-1997.
XX
PF   17-APR-1992;   92US-0872643.
XX
PR   26-JAN-1994;   94US-0188228.
PR   17-APR-1992;   92US-0872643.
PR   19-APR-1993;   93US-0049460.
XX
PA   (DOHE-) DOHENY EYE INST.
XX
PI   Suzuki S;
XX
DR   WPI; 1997-108328/10.
DR   N-PSDB; AAT61920.
XX
PT   Antibodies to cadherin proteins - useful as cadherin antagonists,
PT   etc.
XX
PS   Claim 1; Columns 67-72; 59pp; English.
XX
CC   The present sequence is full length human cadherin-4, which
CC   is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA
CC   was isolated from a foetal brain cDNA library, using probes based
CC   on homologous rat cadherin cDNA.
CC   Antibodies or fragments that specifically bind the human cadherin
CC   can be used to purify the cadherin, determine its tissue expression
CC   and antagonise its ligand/antiligand binding activities.
XX
SQ   Sequence   916 AA;

  Query Match             15.8%;  Score 684.5;  DB 18;  Length 916;
  Best Local Similarity   29.3%;  Pred. No. 2.3e-46;
  Matches  176;  Conservative  106;  Mismatches  288;  Indels   31;  Gaps   12;

Qy     239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
           |  | |: : |||  | | :: ::|    ||   :||:  |  :: |   |||:   | :
Db     171 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFSINSMSGRM 229

Qy     293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
           |||:|:|||| :|   | | |  |  :  |:::::| |:||| |  : |    | |
Db     230 YVTRPMDREEHASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNHPEFINQVYNCSVDEG 289

Qy     353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
            : |  : |:||:| |:  ||| · : |||| |||: |   :| | :  | :         :
Db     290 SKPGTYVMTITANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349

Qy     413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
           :   || : :: :| :              | | |:||    |  | :       : ::
Db     350 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPSEFTASTFAG--EVPENSVE 407

Qy     467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
           : :  :   | |:| : :   :| || ||  |   | ||| || | | : | :|:|
Db     408 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 467

Qy     527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
             :     |  ||  |:: :  | |  |: : |:|||| |  :    :: | |  || :
Db     468 FMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPYFPSNHKLIRLEEGVPPGTVLT 527

Qy     587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
             :| ||:      : ||   |   || |:  |:| :|| ||||:     : |   :|
Db     528 TFSAVDPDRFMQQAVRYSKLSDPASWLHINATNGQITTVAVLDRESLYTKNNVYEATFLA 587

Qy     640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFR 699
           : |    |     : |:|:||| | |      :    |:|   :   | | | |
Db     588 ADNGIPPASGTGTLQIYLIDINDNAPELLPKEAQICERPNLNA---INITAADADVHPNI 644

Qy     700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIV 758
           ||: |       ::: :| ::::|| :|:|| |    |   | | | : | | |||
Db     645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPIIVTDSGNPPLSNTS 704

Qy     759 SLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--FIRIKKDKGK 816
           : |  | | :       |    :   |:  | ::  |: | |:| :|  |:   | : |
Db     705 IIKVKVCPCDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTMVLLFVMWMKRREK 761
```

```
Qy    817 D 817
          :
Db    762 E 762


RESULT 15
AAR63533
ID   AAR63533 standard; Protein; 896 AA.
XX
AC   AAR63533;
XX
DT   07-JUN-1995  (first entry)
XX
DE   Human HT-1376 cell-derived Leukocyte Chemotactic Factor.
XX
KW   rHT-LCF; HT-1376 cell derived leukocyte chemotactic factor;
KW   immunopotentiator; monocyte and macrophage migration factor;
KW   wound healing.
XX
OS   Homo sapiens (cell line HT-1376).
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..30
FT                   /label= signal_peptide
FT   Protein         31..896
FT                   /label= rHT-LCF
XX
PN   WO9421809-A.
XX
PD   29-SEP-1994.
XX
PF   11-MAR-1994;   94WO-JP00397.
XX
PR   15-MAR-1993;   93JP-0054349.
XX
PA   (CYTO-) INST CYTOSIGNAL RES INC.
XX
PI   Kawamura K,  Watanabe K;
XX
DR   WPI; 1994-317029/39.
DR   N-PSDB; AAQ72597.
XX
PT   Monocyte and macrophage migration factor protein - is
PT   immuno-potentiator and vulnerary and is isolated from culture of
PT   HT-1376 cell line by recombinant methods
XX
PS   Claim 5; Page 34-41; 68pp; Japanese.
XX
CC   Sequences coding for proteins having monocyte and macrophage
CC   migration factor activity were isolated from a cDNA library
CC   prepared from human HT-1376 cells (ATCC CRL-1472). The proteins
CC   encoded by the isolated cDNA clones are useful as
CC   immunopotentiators for treatment of infections, immune
CC   deficiency diseases and cancer. The proteins are also useful in
CC   wound healing preparations and for treatment of skin diseases.
XX
SQ   Sequence   896 AA;

  Query Match             15.5%;  Score 671.5;  DB 15;  Length 896;
  Best Local Similarity   30.8%;  Pred. No. 2.5e-45;
  Matches  187;  Conservative  106;  Mismatches  263;  Indels   51;  Gaps   22;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQEG 290
          | || |  | |||  | |: : ||  :|    |::        ||| |  |    |  |
Db    137 W-APIPCSMQENSLGPFPLFLQQVE-SDAAQNYTVFYSISGRGVDKEPLNLFYIERD-TG 193

Qy    291 DIYVTQPLDREEKDAY--VFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFE 348
          ::: |:|:|||| | :  : ||   | |   |  || : ::|:| ||| |     :  ||
Db    194 NLFCTRPVDREEYDVFDLIAYASTADGYSADL--PLPLPIRVEDENDNHPVFTEAIYNFE 251

Qy    349 VQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQ 408
          | |: | | ::| : | |||| :| :: | | |::|||: |  ||| :    |::
Db    252 VLESSRPGTTVGVVCATDRDEPDTMHTRLKYSILQQTPRSP--GLFSVHPSTGVITTVSH 309

Qy    409 SLKKQDTPQYNLTIEVSDKDFK-------TLCFVQINVIDINDQIPIFEKSDYGNLTLAE 461
          | ::   :|:| ::| | | :        : |   | | | ||  | | :: |      |
Db    310 YLDREVVDKYSLIMKVQDMDGQFFGLIGTSTCI--ITVTDSNDNAPTFRQNAYE--AFVE 365

Qy    462 DTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDF 521
          :      || |   | |   | : :: : |:||:  |   : ||  || | : : |||::
Db    366 ENAFNVEILRIPIEDKDLINTANWRVNFTILKGNENGHFKISTDKETNEGVLSVVKPLNY 425

Qy    522 ETAAVSNIVFKAENPEPLVFGV-KYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVA 580
          |     |:     |  |    : :  | : |  |: | |::| |: :      :: |::|
Db    426 EENRQVNLEIGVNNEAPFARDIPRVTALNRALVTVHVRDLDEGPECTPAAQYVRIKENLA 485

Qy    581 IGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREAGSP----Y 633
          :|:|:    | |||   |  : |    | :||: || ::| | :   ||||  :|    |
Db    486 VGSKINGYKAYDPENRNGNGLRYKKLHDPKGWITIDEISGSIITSKILDREVETPKNELY 545

Qy    634 RVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDD 693
          : |:| :    | :      :  : |||||| : ::|  :  | |       ::  | |
Db    546 NITVLAIDKDDRSCTGTLAVN--IEDVNDNPPEILQEY--VVICKPKMGYTDIL--AVDP 599

Qy    694 DQHLFRGPHFTFSLGSGS--LQNDWEVSKINGTHARLS-TRHTDFEERAYVVLIRINDGG 750
          |: :   | | ||| : |  :   | ::|:| | ||||  ::  |:|  | : | : |
Db    600 DEPVHGAP-FYFSLPNTSPEISRLWSLTKVNDTAARLSYQKNAGFQE--YTIPITVKD-- 654

Qy    751 RPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAV-GILLTTLLVIGIILAVVFIR 809
          |        | |  | |   :  |      ||:     |:  |||   |:  ::| :|
Db    655 RAGQAATKLLRVNLCECTHPTQCRATSRSTGVILGKWAILAILLGIALLFSVLLTLVCGV 714

Qy    810 IKKDKGK 816
              |||
Db    715 FGATKGK 721


Search completed: January 15, 2003, 08:17:11
Job time : 67.6849 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:16:02 ; Search time 171.909 Seconds
(without alignments)
96.208 Million cell updates/sec

Title: US-09-079-819-178
Perfect score: 4321
Sequence: 1 MILQAHLHSLCLLMLYLATG.....DKGKDNVESAQASEVKPLRS 832

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 120991 seqs, 19878514 residues

Total number of hits satisfying chosen parameters: 120991

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4321	100.0	832	9	US-10-025-380-1081	Sequence 1081, Ap
2	4321	100.0	832	10	US-09-922-217-1081	Sequence 1081, Ap
3	4321	100.0	832	10	US-09-833-263-1081	Sequence 1081, Ap
4	4208	97.4	958	9	US-10-025-380-1087	Sequence 1087, Ap
5	4208	97.4	958	10	US-09-922-217-1087	Sequence 1087, Ap
6	4208	97.4	958	10	US-09-833-263-1087	Sequence 1087, Ap
7	908.5	21.0	807	9	US-10-063-547-98	Sequence 98, Appl
8	908.5	21.0	807	12	US-10-006-867-98	Sequence 98, Appl
9	877.5	20.3	830	10	US-09-838-529-5	Sequence 5, Appli
10	699.5	16.2	906	10	US-09-905-983-46	Sequence 46, Appl
11	696	16.1	912	10	US-09-905-983-2	Sequence 2, Appli
12	690.5	16.0	906	10	US-09-746-491-46	Sequence 46, Appl
13	684	15.8	912	10	US-09-746-491-49	Sequence 49, Appl
14	679.5	15.7	906	10	US-09-746-491-48	Sequence 48, Appl
15	676.5	15.7	877	10	US-09-746-491-47	Sequence 47, Appl
16	604	14.0	899	10	US-09-905-983-5	Sequence 5, Appli
17	593	13.7	878	9	US-10-165-049-2	Sequence 2, Appli
18	593	13.7	878	9	US-10-165-049-3	Sequence 3, Appli
19	593	13.7	878	10	US-09-905-983-48	Sequence 48, Appl
20	576	13.3	3014	10	US-09-737-149-2	Sequence 2, Appli
21	571	13.2	3034	10	US-09-737-149-25	Sequence 25, Appl
22	571	13.2	3034	10	US-09-737-149-30	Sequence 30, Appl
23	564.5	13.1	788	12	US-10-003-152-6	Sequence 6, Appli
24	558	12.9	3313	10	US-09-737-149-29	Sequence 29, Appl
25	552	12.8	3503	9	US-10-108-605-237	Sequence 237, App
26	549	12.7	781	10	US-09-860-868-2	Sequence 2, Appli
27	543.5	12.6	693	10	US-09-919-497-55	Sequence 55, Appl
28	537	12.4	1274	10	US-09-746-491-12	Sequence 12, Appl
29	530	12.3	2923	10	US-09-788-711A-4	Sequence 4, Appli
30	530	12.3	2956	10	US-09-788-711A-2	Sequence 2, Appli
31	526.5	12.2	829	10	US-09-905-983-50	Sequence 50, Appl
32	512.5	11.9	413	10	US-09-746-491-43	Sequence 43, Appl
33	484.5	11.2	784	10	US-09-905-983-52	Sequence 52, Appl
34	453	10.5	772	9	US-09-978-295A-264	Sequence 264, App
35	453	10.5	772	9	US-09-978-697-264	Sequence 264, App
36	453	10.5	772	9	US-09-978-192A-264	Sequence 264, App
37	453	10.5	772	9	US-10-066-500-121	Sequence 121, App
38	453	10.5	772	9	US-09-999-832A-264	Sequence 264, App
39	453	10.5	772	9	US-09-978-189-264	Sequence 264, App
40	453	10.5	772	9	US-10-174-590-106	Sequence 106, App
41	453	10.5	772	9	US-10-176-758-106	Sequence 106, App
42	453	10.5	772	12	US-10-052-586-106	Sequence 106, App
43	428	9.9	678	9	US-09-908-193-24	Sequence 24, Appl
44	423	9.8	1713	10	US-09-737-149-27	Sequence 27, Appl
45	419	9.7	724	9	US-09-908-193-25	Sequence 25, Appl

ALIGNMENTS

RESULT 1
US-10-025-380-1081
; Sequence 1081, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C14
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1081
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1081

Query Match 100.0%; Score 4321; DB 9; Length 832;
Best Local Similarity 100.0%; Pred. No. 4.1e-316;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
```

```
Qy      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
```

RESULT 2
US-09-922-217-1081
; Sequence 1081, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922,217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1081
; LENGTH: 832
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1081

Query Match 100.0%; Score 4321; DB 10; Length 832;
Best Local Similarity 100.0%; Pred. No. 4.1e-316;
Matches 832; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
```

RESULT 3
US-09-833-263-1081

```
; Sequence 1081, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
;  CURRENT FILING DATE:  2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1081
;   LENGTH: 832
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-263-1081

  Query Match             100.0%;  Score 4321;  DB 10;  Length 832;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-316;
  Matches  832;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy     61 ELTGETDNIFVIEREGLLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ELTGETDNIFVIEREGLLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy    121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy    181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy    241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy    301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy    361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy    421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy    721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832


RESULT 4
US-10-025-380-1087
; Sequence 1087, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole L.
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Skeiky, Yasir A. W.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick Thomas S.
;  APPLICANT:  Carter, Darrick
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C14
;  CURRENT APPLICATION NUMBER: US/10/025,380
;  CURRENT FILING DATE:  2001-12-19
;  NUMBER OF SEQ ID NOS: 1129
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1087
;   LENGTH: 958
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-025-380-1087

  Query Match             97.4%;  Score 4208;  DB 9;  Length 958;
  Best Local Similarity   99.4%;  Pred. No. 1.5e-307;
  Matches  813;  Conservative    1;  Mismatches    2;  Indels    2;  Gaps    1;

Qy     17 LATG--YGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIER 74
          ||:|    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 LASGSAAAQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIER 200

Qy     75 EGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQSKYEGS 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 EGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQSKYEGS 260

Qy    135 VRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTRE 194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 VRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTRE 320

Qy    195 GSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 380

Qy    255 HPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKD 314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 HPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKD 440

Qy    315 EYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTAN 374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 EYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTAN 500

Qy    375 SFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 SFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 560
```

```
Qy    435 VQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKG 494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 VQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKG 620

Qy    495 DSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFT 554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 DSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFT 680

Qy    555 LIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 LIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 740

Qy    615 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGL 674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    741 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGL 800

Qy    675 FFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 734
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    801 FFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 860

Qy    735 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLT 794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    861 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLT 920

Qy    795 TLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
          |||||||||||||||||||||||||||||||||||||
Db    921 TLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 958


RESULT 5
US-09-922-217-1087
; Sequence 1087, Application US/09922217
; Patent No. US20020076414A1
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Lodes, Michael J.
;  APPLICANT:  Secrist, Heather
;  APPLICANT:  Benson, Darin R.
;  APPLICANT:  Meagher, Madeleine Joy
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Jiang, Yuqiu
;  APPLICANT:  Smith, Carole Lynn
;  APPLICANT:  King, Gordon E.
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
;  TITLE OF INVENTION:  OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C13
;  CURRENT APPLICATION NUMBER: US/09/922,217
;  CURRENT FILING DATE:  2001-08-03
;  NUMBER OF SEQ ID NOS: 1124
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1087
;   LENGTH: 958
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-922-217-1087

  Query Match             97.4%;  Score 4208;  DB 10;  Length 958;
  Best Local Similarity   99.4%;  Pred. No. 1.5e-307;
  Matches  813;  Conservative    1;  Mismatches    2;  Indels    2;  Gaps    1;

Qy     17 LATG--YGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIER 74
          ||:|    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 LASGSAAAQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIER 200

Qy     75 EGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQSKYEGS 134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 EGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQSKYEGS 260

Qy    135 VRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTRE 194
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 VRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTRE 320

Qy    195 GSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 254
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321 GSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 380

Qy    255 HPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKD 314
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381 HPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKD 440

Qy    315 EYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTAN 374
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441 EYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTAN 500

Qy    375 SFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 434
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501 SFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 560

Qy    435 VQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKG 494
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561 VQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKG 620

Qy    495 DSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFT 554
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621 DSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFT 680

Qy    555 LIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 614
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681 LIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 740

Qy    615 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGL 674
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    741 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGL 800

Qy    675 FFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 734
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    801 FFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 860

Qy    735 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLT 794
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    861 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLT 920

Qy    795 TLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
          |||||||||||||||||||||||||||||||||||||
Db    921 TLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 958


RESULT 6
US-09-833-263-1087
; Sequence 1087, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Aijun
;  APPLICANT:  Clapper, Jonathan D.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Meagher, Madeleine J.
;  TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
;  FILE REFERENCE: 210121.471C12
;  CURRENT APPLICATION NUMBER: US/09/833,263
;  CURRENT FILING DATE:  2001-04-10
;  NUMBER OF SEQ ID NOS: 1093
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1087
;   LENGTH: 958
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-833-263-1087

  Query Match             97.4%;  Score 4208;  DB 10;  Length 958;
  Best Local Similarity   99.4%;  Pred. No. 1.5e-307;
  Matches  813;  Conservative    1;  Mismatches    2;  Indels    2;  Gaps    1;
```

```
Qy      17 LATG--YGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIER 74
           ||:|    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 LASGSAAAQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIER 200

Qy      75 EGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQSKYEGS 134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 EGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQSKYEGS 260

Qy     135 VRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTRE 194
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 VRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTRE 320

Qy     195 GSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 254
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDP 380

Qy     255 HPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKD 314
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     381 HPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKD 440

Qy     315 EYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTAN 374
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     441 EYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTAN 500

Qy     375 SFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 434
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     501 SFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCF 560

Qy     435 VQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKG 494
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     561 VQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKG 620

Qy     495 DSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFT 554
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     621 DSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFT 680

Qy     555 LIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 614
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     681 LIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDH 740

Qy     615 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGL 674
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     741 VTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGL 800

Qy     675 FFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 734
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     801 FFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 860

Qy     735 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLT 794
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     861 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLT 920

Qy     795 TLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           ||||||||||||||||||||||||||||||||||||||
Db     921 TLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 958


RESULT 7
US-10-063-547-98
; Sequence 98, Application US/10063547
; Publication No. US20020182638A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/063,547
;  CURRENT FILING DATE:  2002-05-02
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 98
;   LENGTH: 807
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-063-547-98

  Query Match             21.0%;  Score 908.5;  DB 9;  Length 807;
  Best Local Similarity   30.2%;  Pred. No. 3.9e-60;
  Matches  237;  Conservative  159;  Mismatches  335;  Indels   53;  Gaps   21;

Qy      62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:    |:  | :: : | |   ||||||:: : |||     :| :: || |: : |
Db      57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEQAEYQLQVTLEMQDGHVLWGPQPVLVHV 116

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : : :||| |||:: |:| |:| | |  | : |: | |   :
Db     117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRFHILSQAPAQPS 176

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
             ||:  : ||::|: :||  |: |   :| |:: ||||| |: :    | :|:: : |
Db     177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKDMGDQA-SGHQATATVEVSIIE 235

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           : | : :|: : ||    :|  : || |:     | |   |  | |  || :: ||::|||:
Db     236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFEVNAEGNLYVTR 292

Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
           |||| :  |:    |:: :|:  : |||:|| | | ||| | ||       : |    |
Db     293 ELDREAQAEYLLQVRAQNSHGEDYAAPLELHVLVMDENDNVPICPPRDPTVSIPELSPPG 352

Qy     357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQLAKQSLKK-QD 414
           :  |:| | |   : || : |:::   |:  ::|  | :   :| : |    |:  |:
Db     353 TEVTRLSAEDADAPGSPNSHVVYQLLSPEPEDGVEGRAFQVDPTSGSVTLGVLPLRAGQN 412

Qy     415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTI 472
           | :::  :  | : | |:: | |||  | |  |  | ::| ||   |: :  :
Db     413 ILLLVLAMDLAGAEGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472

Qy     473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIK--KPLDFETAAVSN 528
           | ||| ||   : ::: : | :||:||  |:| :|  ::|:| ::  | | :| |
Db     473 TAIDADLEP---AFRLMDFAIERGDTEGTFGLDWEP--DSGHVRLRLCKNLSYEAAPSHE 527

Qy     529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
           :|   ::   || |      : |  |::|  |   |:  |  ::| |      |: :  :
Db     528 VVVVVQSVAKLV-GPGPGPGATATVTVLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586

Qy     589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647
              ||    : :||  |: ||| |:  :||: :   |   : |  | | | | :
Db     587 QPSDPISRTLRFSLVNDSEGWLCIEKFSGEVHTAQSLQGAQPGDTYTVLVEAQDTA---- 642

Qy     648 SSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF-EATDDDQHLFRGPHFTFS 706
           | |  |              : | |    | ::   : | |     || ::|:
Db     643 -------LTLAPVPSQ----------YLCTPRQDHGLIVSGPSKDPDLASGHGP-YSFT 683

Qy     707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
           || : ::| || :  :||:|| |:      | | ::: : ::   :      : : |  |
Db     684 LGPNPTVQRDWRLQTLNGSHAYLTLALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740

Qy     766 SC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQA 824
           | ||| | |  |   |:||   |||||: ||: ||| | ::|      : ||  : | :
Db     741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGTLVAIGIFLILIFTHWTMSRKKDPDQPADS 800

Qy     825 SEVK 828
             :|
Db     801 VPLK 804
```

```
RESULT 8
US-10-006-867-98
; Sequence 98, Application US/10006867
; Patent No. US20020119130A1
; GENERAL INFORMATION:
;  APPLICANT: Eaton,Dan L.
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Grimaldi,Christopher J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Watanabe,Colin K.
;  APPLICANT:  Wood,William I.
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3230R1C1
;  CURRENT APPLICATION NUMBER: US/10/006,867
;  CURRENT FILING DATE:  2001-12-06
;  PRIOR APPLICATION NUMBER: 60/063435
;  PRIOR FILING DATE: 1997-10-29
;  PRIOR APPLICATION NUMBER: 60/064215
;  PRIOR FILING DATE: 1997-10-29
;  PRIOR APPLICATION NUMBER: 60/082797
;  PRIOR FILING DATE: 1998-04-22
;  PRIOR APPLICATION NUMBER: 60/083495
;  PRIOR FILING DATE: 1998-04-29
;  PRIOR APPLICATION NUMBER: 60/085579
;  PRIOR FILING DATE: 1998-05-15
;  PRIOR APPLICATION NUMBER: 60/087759
;  PRIOR FILING DATE: 1998-06-02
;  PRIOR APPLICATION NUMBER: 60/088021
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088029
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088030
;  PRIOR FILING DATE: 1998-06-04
;  PRIOR APPLICATION NUMBER: 60/088734
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088740
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088811
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088824
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088825
;  PRIOR FILING DATE: 1998-06-10
;  PRIOR APPLICATION NUMBER: 60/088863
;  PRIOR FILING DATE: 1998-06-11
;  PRIOR APPLICATION NUMBER: 60/089105
;  PRIOR FILING DATE: 1998-06-12
;  PRIOR APPLICATION NUMBER: 60/089514
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/089653
;  PRIOR FILING DATE: 1998-06-17
;  PRIOR APPLICATION NUMBER: 60/089952
;  PRIOR FILING DATE: 1998-06-19
;  PRIOR APPLICATION NUMBER: 60/090246
;  PRIOR FILING DATE: 1998-06-22
;  PRIOR APPLICATION NUMBER: 60/090444
;  PRIOR FILING DATE: 1998-06-24
;  PRIOR APPLICATION NUMBER: 60/090688
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090696
;  PRIOR FILING DATE: 1998-06-25
;  PRIOR APPLICATION NUMBER: 60/090862
;  PRIOR FILING DATE: 1998-06-26
;  PRIOR APPLICATION NUMBER: 60/091628
;  PRIOR FILING DATE: 1998-07-02
;  PRIOR APPLICATION NUMBER: 60/096012
;  PRIOR FILING DATE: 1998-08-10
;  PRIOR APPLICATION NUMBER: 60/096757
;  PRIOR FILING DATE: 1998-08-17
;  PRIOR APPLICATION NUMBER: 60/096949
;  PRIOR FILING DATE: 1998-08-18
;  PRIOR APPLICATION NUMBER: 60/096959
;  PRIOR FILING DATE: 1998-08-18
;  PRIOR APPLICATION NUMBER: 60/097954
;  PRIOR FILING DATE: 1998-08-26
;  PRIOR APPLICATION NUMBER: 60/097971
;  PRIOR FILING DATE: 1998-08-26
;  PRIOR APPLICATION NUMBER: 60/097979
;  PRIOR FILING DATE: 1998-08-26
;  PRIOR APPLICATION NUMBER: 60/098749
;  PRIOR FILING DATE: 1998-09-01
;  PRIOR APPLICATION NUMBER: 60/099741
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099763
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099792
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099812
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/099815
;  PRIOR FILING DATE: 1998-09-10
;  PRIOR APPLICATION NUMBER: 60/100627
;  PRIOR FILING DATE: 1998-09-16
;  PRIOR APPLICATION NUMBER: 60/100662
;  PRIOR FILING DATE: 1998-09-16
;  PRIOR APPLICATION NUMBER: 60/100683
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/100684
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/100930
;  PRIOR FILING DATE: 1998-09-17
;  PRIOR APPLICATION NUMBER: 60/101279
;  PRIOR FILING DATE: 1998-09-22
;  PRIOR APPLICATION NUMBER: 60/101475
;  PRIOR FILING DATE: 1998-09-23
;  PRIOR APPLICATION NUMBER: 60/101738
;  PRIOR FILING DATE: 1998-09-24
;  PRIOR APPLICATION NUMBER: 60/101743
;  PRIOR FILING DATE: 1998-09-24
;  PRIOR APPLICATION NUMBER: 60/101916
;  PRIOR FILING DATE: 1998-09-24
;  PRIOR APPLICATION NUMBER: 60/102570
;  PRIOR FILING DATE: 1998-09-30
;  PRIOR APPLICATION NUMBER: 60/103449
;  PRIOR FILING DATE: 1998-10-06
;  PRIOR APPLICATION NUMBER: 60/103678
;  PRIOR FILING DATE: 1998-10-08
;  PRIOR APPLICATION NUMBER: 60/103679
;  PRIOR FILING DATE: 1998-10-08
;  PRIOR APPLICATION NUMBER: 60/103711
;  PRIOR FILING DATE: 1998-10-08
;  PRIOR APPLICATION NUMBER: 60/105000
;  PRIOR FILING DATE: 1998-10-20
;  PRIOR APPLICATION NUMBER: 60/105002
;  PRIOR FILING DATE: 1998-10-20
;  PRIOR APPLICATION NUMBER: 60/105881
;  PRIOR FILING DATE: 1998-10-27
;  PRIOR APPLICATION NUMBER: 60/106030
;  PRIOR FILING DATE: 1998-10-28
;  PRIOR APPLICATION NUMBER: 60/106464
;  PRIOR FILING DATE: 1998-10-30
;  PRIOR APPLICATION NUMBER: 60/106856
;  PRIOR FILING DATE: 1998-11-03
;  PRIOR APPLICATION NUMBER: 60/108807
;  PRIOR FILING DATE: 1998-11-17
;  PRIOR APPLICATION NUMBER: 60/112419
;  PRIOR FILING DATE: 1998-12-15
;  PRIOR APPLICATION NUMBER: 60/112422
;  PRIOR FILING DATE: 1998-12-15
;  PRIOR APPLICATION NUMBER: 60/112853
;  PRIOR FILING DATE: 1998-12-16
;  PRIOR APPLICATION NUMBER: 60/113011
```

```
;  PRIOR FILING DATE: 1998-12-16
;  PRIOR APPLICATION NUMBER: 60/112854
;  PRIOR FILING DATE: 1998-12-16
;  PRIOR APPLICATION NUMBER: 60/113300
;  PRIOR FILING DATE: 1998-12-22
;  PRIOR APPLICATION NUMBER: 60/113408
;  PRIOR FILING DATE: 1998-12-22
;  PRIOR APPLICATION NUMBER: 60/113430
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/113621
;  PRIOR FILING DATE: 1998-12-23
;  PRIOR APPLICATION NUMBER: 60/114223
;  PRIOR FILING DATE: 1998-12-30
;  PRIOR APPLICATION NUMBER: 60/115614
;  PRIOR FILING DATE: 1999-01-12
;  PRIOR APPLICATION NUMBER: 60/116527
;  PRIOR FILING DATE: 1999-01-20
;  PRIOR APPLICATION NUMBER: 60/116843
;  PRIOR FILING DATE: 1999-01-22
;  PRIOR APPLICATION NUMBER: 60/119285
;  PRIOR FILING DATE: 1999-02-09
;  PRIOR APPLICATION NUMBER: 60/119287
;  PRIOR FILING DATE: 1999-02-09
;  PRIOR APPLICATION NUMBER: 60/119525
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: 60/119549
;  PRIOR FILING DATE: 1999-02-10
;  PRIOR APPLICATION NUMBER: 60/120014
;  PRIOR FILING DATE: 1999-02-11
;  PRIOR APPLICATION NUMBER: 60/129122
;  PRIOR FILING DATE: 1999-04-13
;  PRIOR APPLICATION NUMBER: 60/129674
;  PRIOR FILING DATE: 1999-04-16
;  PRIOR APPLICATION NUMBER: 60/131291
;  PRIOR FILING DATE: 1999-04-27
;  PRIOR APPLICATION NUMBER: 60/138387
;  PRIOR FILING DATE: 1999-06-09
;  PRIOR APPLICATION NUMBER: 60/144791
;  PRIOR FILING DATE: 1999-07-20
;  PRIOR APPLICATION NUMBER: 60/169495
;  PRIOR FILING DATE: 1999-12-07
;  PRIOR APPLICATION NUMBER: 60/175481
;  PRIOR FILING DATE: 2000-01-11
;  PRIOR APPLICATION NUMBER: 60/191007
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/199397
;  PRIOR FILING DATE: 2000-04-25
;  PRIOR APPLICATION NUMBER: 09/380139
;  PRIOR FILING DATE: 1998-08-25
;  PRIOR APPLICATION NUMBER: 09/311832
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 09/380137
;  PRIOR FILING DATE: 1999-08-25
;  PRIOR APPLICATION NUMBER: 09/380138
;  PRIOR FILING DATE: 1999-08-25
;  PRIOR APPLICATION NUMBER: 09/380142

  Query Match             21.0%;  Score 908.5;  DB 12;  Length 807;
  Best Local Similarity   30.2%;  Pred. No. 3.9e-60;
  Matches  237;  Conservative  159;  Mismatches  335;  Indels   53;  Gaps   21;

Qy      62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:    |:  | :: : | |   ||||||  :: : |||     :| :: || |: : |
Db      57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEQAEYQLQVTLEMQDGHVLWGPQPVLVHV 116

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : : :||| |||:: |:| |:| | |  | : |: | |   :
Db     117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRFHILSQAPAQPS 176

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
              ||:  : ||::|: :||  |: |   :| |:: ||||| |: :    | :|:: : |
Db     177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKDMGDQA-SGHQATATVEVSIIE 235

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           : | : :|: : ||    :|  : || |:     | |   |  |  || :: ||::|||:
Db     236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFEVNAEGNLYVTR 292

Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
           |||| :  |:    |:: :|:  : |||:|| | | ||| | ||       : |    |
Db     293 ELDREAQAEYLLQVRAQNSHGEDYAAPLELHVLVMDENDNVPICPPRDPTVSIPELSPPG 352

Qy     357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQLAKQSLKK-QD 414
             :  |:| | |   : || : |:::   |:  ::|  | :   :| : |    |:  |:
Db     353 TEVTRLSAEDADAPGSPNSHVVYQLLSPEPEDGVEGRAFQVDPTSGSVTLGVLPLRAGQN 412

Qy     415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTI 472
             | :::  :  | : | |:: | ||||  | |  |  | ::| ||   |: :  :
Db     413 ILLLVLAMDLAGAEGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472

Qy     473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIK--KPLDFETAAVSN 528
           | ||| ||   : ::: : | :||:||  |:| :|  ::|:| ::  | | :| |
Db     473 TAIDADLEP---AFRLMDFAIERGDTEGTFGLDWEP--DSGHVRLRLCKNLSYEAAPSHE 527

Qy     529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
           :|   ::   || |      : |  |::|  |   |:  |  ::| |      |: :  :
Db     528 VVVVVQSVAKLV-GPGPGPGATATVTVLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586

Qy     589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647
             ||    : :||  |: ||| |:  :||: :   |   : |  | | | | :
Db     587 QPSDPISRTLRFSLVNDSEGWLCIEKFSGEVHTAQSLQGAQPGDTYTVLVEAQDTA---- 642

Qy     648 SSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF-EATDDDQHLFRGPHFTFS 706
             | |  |               : | |    | ::   : | |     || ::|:
Db     643 -------LTLAPVPSQ----------YLCTPRQDHGLIVSGPSKDPDLASGHGP-YSFT 683

Qy     707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
           || : ::| || :  :||:|| |:      | | ::: : ::   :      : : |  |
Db     684 LGPNPTVQRDWRLQTLNGSHAYLTLALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 740

Qy     766 SC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQA 824
             | ||| | |  |   |:||   |||||: ||: ||| | ::|      : ||  : | :
Db     741 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGTLVAIGIFLILIFTHWTMSRKKDPDQPADS 800

Qy     825 SEVK 828
             :|
Db     801 VPLK 804


RESULT 9
US-09-838-529-5
; Sequence 5, Application US/09838529
; Patent No. US20020081599A1
; GENERAL INFORMATION:
;  APPLICANT: Curtis, Rory A. J.
;  TITLE OF INVENTION: 57809 AND 57798, NOVEL HUMAN CADHERIN
;  TITLE OF INVENTION:  MOLECULES AND USES THEREFOR
;  FILE REFERENCE: MNI-145
;  CURRENT APPLICATION NUMBER: US/09/838,529
;  CURRENT FILING DATE:  2001-04-18
;  PRIOR APPLICATION NUMBER: 60/198,466
;  PRIOR FILING DATE: 2000-04-18
;  NUMBER OF SEQ ID NOS: 7
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 830
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-838-529-5

  Query Match             20.3%;  Score 877.5;  DB 10;  Length 830;
  Best Local Similarity   30.5%;  Pred. No. 8.6e-58;
  Matches  245;  Conservative  141;  Mismatches  353;  Indels   63;  Gaps   22;

Qy      62 LTGETD----NIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:::    | | :: : | |    | |||| :: : |||     :| |: ||  :|: |
Db      59 LSGDSNTADQNTFAVDTDSGFLVATRTLDREEKAEYQLQVTLESEDGRILWGPQLVTVHV 118
```

```
Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
Db    119 KDENDQVPQFSQAIYRAQLSQGTRPGVPFLFLEASDGDAPGTANSDLRFHILSQSPPQPL 178

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
Db    179 PDMFQLDPHLGALALSPSGSTSLDHALEETYQLLVQVKDMGDQPSGHQAIAT-VEISIVE 237

Qy    237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
Db    238 NSWAPLEPVHLAENLKVVYPHSIAQVHWSGGDVHYQL---ESQPPGPFDVDTEGMLHVTM 294

Qy    297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
Db    295 ELDREAQAEYQLQVRAQNSHGEDYAEPLELQVVVMDENDNAPVCSPHDPTVNIPELSPPG 354

Qy    357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQL------AKQS 409
Db    355 TEIARLSAEDLDAPGSPNSHIVYQLLSPEPEEGAENKAFELDPTSGSVTLGTAPLHAGQS 414

Qy    410 LKKQDTPQYNLTIEV--SDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGS 467
Db    415 ILLQ-----VLAVDLAGSESGLSSTCEVTVMVTDVNNHAPEFINSQIGPVTLPEDVKPGA 469

Qy    468 TILTIQATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAA 525
Db    470 LVATLMATDADLEP---AFRLMDFAIEEGDPEGIFDLSWEPDSDHVQLRLRKNLSYEAAP 526

Qy    526 VSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKV 585
Db    527 DHKVVVVVSNIEELV-GPGPGPAATATVTILVERVVAPLKLDQESYETSIPVSTPAGSLL 585

Qy    586 GNVTAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVA--TEV 642
Db    586 LTIQPSDPMSRTLRFSLVNDSEGWLCIKEVSGEVHTAQSLQGAQPGDTYTVLVEAQDTDK 645

Qy    643 GGSSLSSVSEFHLI------LMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQH 696
Db    646 PGLSTSATVVIHFLKASPVPALTLSAGPSR--------HLCTPRQDYGVVVSGVSEDPDL 697

Qy    697 LFRGPHFTFSLG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRI-NDGGRPPL 754
Db    698 ANRNGPYSFALGPNPTVQRDWRLQPLNDSHAYLTLALHWVEPGEYMVPVVVHHDTHMWQL 757

Qy    755 EGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRI--- 810
Db    758 Q----VKVIVCRCNVEGQCMRKVGRMKGMPTKLSAVGVLLGTLAAIGFILILVFTHLALA 813

Qy    811 KKDKGKDNVESAQASEVKPLRS 832
Db    814 RKDLD-------QPADSVPLKA 828
```

RESULT 10
US-09-905-983-46
; Sequence 46, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-905-983-46

Query Match 16.2%; Score 699.5; DB 10; Length 906;
Best Local Similarity 27.4%; Pred. No. 2.2e-44;
Matches 210; Conservative 122; Mismatches 331; Indels 103; Gaps 21;

```
Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
Db     47 KDVHEGQP-LLNVKF-----SNCNGKRKVQYESSEPADFKVDEDGMVY--AVRSFPLSSE 98

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
Db     99 HAKFLIY-----------AQDKETQEKWQVAVKLSLKPT--LTEESVKESAEVEEIVFP 144

Qy    235 -----------TENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQ 270
Db    145 RQFSKHSGHLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYTVTGPGA- 202

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
Db    203 ------DQPPTGIFIINPISGQLSVTKPLDREQIARFHLRAHAVDINGNQVENPIDIVIN 256

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
Db    257 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
Db    317 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGIPTYGLSNTATAVITVTDVN 376

Qy    444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502
Db    377 DNPPEFTAMTFYGEV---PENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAI 433

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
Db    434 QTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 493

Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
Db    494 NPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQI 553

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675
Db    554 TTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL------- 606

Qy    676 FCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727
Db    607 ---PQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQ 663

Qy    728 LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV- 785
Db    664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIV 716

Qy    786 --GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
Db    717 GAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762
```

RESULT 11
US-09-905-983-2
; Sequence 2, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Einat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2

```
;   LENGTH: 912
;   TYPE: PRT
;   ORGANISM: Gallus gallus
US-09-905-983-2

  Query Match             16.1%;  Score 696;  DB 10;  Length 912;
  Best Local Similarity   26.9%;  Pred. No. 4.1e-44;
  Matches  209;  Conservative  117;  Mismatches  328;  Indels  124;  Gaps   22;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183
Db     41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy    184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENSFSDTT 228
Db     82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy    229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269
Db    137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy    270 QYSLVDKEKLPRFPFSI-----DQ-----------EGDIYVTQPLDREEKDAYVFYAVAK 313
Db    190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy    314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373
Db    246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy    374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428
Db    306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy    429 -FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
Db    366 GLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWN 422

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
Db    423 ARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISYSL 602
Db    483 PQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRYSK 542

Qy    603 RGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILM 658
Db    543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIYLL 602

Qy    659 DVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQND 715
Db    603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657

Qy    716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFR 774
Db    658 WTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC-- 715

Qy    775 PAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
Db    716 -----TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768


RESULT 12
US-09-746-491-46
; Sequence 46, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
;  APPLICANT: Burgess, Catherine E.
;  TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 15966-621
;  CURRENT APPLICATION NUMBER: US/09/746,491
;  CURRENT FILING DATE:  2000-12-20
;  PRIOR APPLICATION NUMBER: USSN 60/171,329
;  PRIOR FILING DATE: 1999-12-21
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 46
;   LENGTH: 906
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-746-491-46

  Query Match             16.0%;  Score 690.5;  DB 10;  Length 906;
  Best Local Similarity   27.5%;  Pred. No. 1e-43;
  Matches  211;  Conservative  124;  Mismatches  326;  Indels  107;  Gaps   23;

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
Db     47 KDVHEGQP-LLNVKF-----SNCNGKRKVQYESSEPADFKVDEDGMVY--AVRSFPLSSE 98

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
Db     99 HAKFLI------VAQDKETQEKWQVA------VKLSLKPT--LTEESVKESAEVEEIVFP 144

Qy    235 ------------TENIWKAPKPVEMVENSTDPHPIKITQVRWN------------DPGAQ 270
Db    145 RQFSKHSGHLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA- 202

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
Db    203 ------DQPPTGIFIINPISGQLSVTKPLDREQIARFHLRAHAVDINGNQVENPIDIVIN 256

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
Db    257 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
Db    317 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDWN 376

Qy    444 DQIPIFEKSD---YGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRL 500
Db    377 DNPP--EETAMTFYGEV---PENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRF 431

Qy    501 GVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDV 560
Db    432 AIQTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDV 491

Qy    561 NEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTG 617
Db    492 NENPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNG 551

Qy    618 EIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTG 673
Db    552 QITTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL----- 606

Qy    674 LFFCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTH 725
Db    607 -----PQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDF 661

Qy    726 ARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPT 784
Db    662 AQLNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDR 714

Qy    785 V---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
Db    715 IVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762


RESULT 13
US-09-746-491-49
; Sequence 49, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
;  APPLICANT: Burgess, Catherine E.
;  TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
```

```
;  FILE REFERENCE: 15966-621
;  CURRENT APPLICATION NUMBER: US/09/746,491
;  CURRENT FILING DATE:  2000-12-20
;  PRIOR APPLICATION NUMBER: USSN 60/171,329
;  PRIOR FILING DATE: 1999-12-21
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
;   LENGTH: 912
;   TYPE: PRT
;   ORGANISM: Gallus gallus
US-09-746-491-49

  Query Match             15.8%;  Score 684;  DB 10;  Length 912;
  Best Local Similarity   26.8%;  Pred. No. 3.2e-43;
  Matches  209;  Conservative  117;  Mismatches  326;  Indels  128;  Gaps   23;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183
Db     41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy    184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENSFSDTT 228
Db     82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy    229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269
Db    137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy    270 QYSLVDKEKLPRFPFSI-----DQ-----------EGDIYVTQPLDREEKDAYVFYAVAK 313
Db    190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy    314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373
Db    246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy    374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428
Db    306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy    429 -FKTLCFVQINVIDINDQIPIFEKSD---YGNLTLAEDTNIGSTILTIQATDADEPFTGS 484
Db    366 GLSNTATAVITVTDWNDNPP--EETAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPA 420

Qy    485 SKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVK 544
Db    421 WNARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQ 480

Qy    545 YNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISY 600
Db    481 HPPQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRY 540

Qy    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLI 656
Db    541 SKLSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIY 600

Qy    657 LMDVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQ 713
Db    601 LLDINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIK 655

Qy    714 NDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSC 772
Db    656 RNWTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC 715

Qy    773 FRPAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
Db    716 -------TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768


RESULT 14
US-09-746-491-48
; Sequence 48, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
;  APPLICANT: Burgess, Catherine E.
;  TITLE OF INVENTION: No. US20020137202A1el Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 15966-621
;  CURRENT APPLICATION NUMBER: US/09/746,491
;  CURRENT FILING DATE:  2000-12-20
;  PRIOR APPLICATION NUMBER: USSN 60/171,329
;  PRIOR FILING DATE: 1999-12-21
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48
;   LENGTH: 906
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-746-491-48

  Query Match             15.7%;  Score 679.5;  DB 10;  Length 906;
  Best Local Similarity   27.5%;  Pred. No. 7e-43;
  Matches  211;  Conservative  122;  Mismatches  330;  Indels  103;  Gaps   24;

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
Db     47 KDVHEGQP-LLNVKFSNCNRK-----RKVQYESSEPADFKVDEDGTVY--AVRSFPLTAF 98

Qy    177 VMYFQI---NNKTG-----AISLTREGSQELNPAKNP-SYNLVISVKDM---GGQSENSF 224
Db     99 QAKFLIYAQDKETQEKWQVAVNLSREPTLTEEPMKEPHEIEEIVFPRQLAKHSGALQRQK 158

Qy    225 SDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIKITQVRWN------------DPGAQYS 272
Db    159 RD------------WVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA--- 202

Qy    273 LVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVK 331
Db    203 ----DQPPTGIFIINPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVINVI 258

Qy    332 DINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMD 391
Db    259 DMNDNRPEFLHQVWNGSVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTPSP 318

Qy    392 GLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQ 445
Db    319 NMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDWNDN 378

Qy    446 IPIFEKSD---YGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502
Db    379 PP--EETAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWNAAYRISGGDPTGRFAI 433

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
Db    434 LTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 493

Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
Db    494 NPYFAPNPKIIRQEEGLHAGTMLTTLTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQI 553

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675
Db    554 TTIAVLDRESPYVQNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL------- 606

Qy    676 FCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727
Db    607 ---PQEAETCETPEPNSINIAALDYDIDPNAGPFAFDLPLSPVTIKRNWTINRLNGDFAQ 663

Qy    728 LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV- 785
Db    664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIV 716

Qy    786 --GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
Db    717 GAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762
```

```
RESULT 15
US-09-746-491-47
; Sequence 47, Application US/09746491
; Patent No. US20020137202A1
; GENERAL INFORMATION:
;  APPLICANT: Burgess, Catherine E.
;  TITLE OF INVENTION: No. US20020137202Alel Proteins and Nucleic Acids Encoding Same
;  FILE REFERENCE: 15966-621
;  CURRENT APPLICATION NUMBER: US/09/746,491
;  CURRENT FILING DATE:  2000-12-20
;  PRIOR APPLICATION NUMBER: USSN 60/171,329
;  PRIOR FILING DATE: 1999-12-21
;  NUMBER OF SEQ ID NOS: 72
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
;   LENGTH: 877
;   TYPE: PRT
;   ORGANISM: Bos taurus
US-09-746-491-47

  Query Match             15.7%;  Score 676.5;  DB 10;  Length 877;
  Best Local Similarity   26.7%;  Pred. No. 1.1e-42;
  Matches  203;  Conservative  126;  Mismatches  341;  Indels   91;  Gaps   20;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLP---------MINN 176
Db      7 FPEDVYSAVLSRDVLEGQPLLNVKFSNC------NGKRKVQYESSEPADFKVDEDGMVYA 60

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
Db     61 VRSFPLSSEHSKFLIYAQDKE-----TQEKWQVAVKLSLKPALPEDSVKESREIEEIVFP 115

Qy    235 ------------TENIWKAPKPVEMVENSTDPHPIKITQVRWN------------DPGAQ 270
Db    116 RQVTKHNGYLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA- 173

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
Db    174 ------DQPPTGIFIINPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVIN 227

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
Db    228 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTP 287

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
Db    288 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDWN 347

Qy    444 DQIPIFEKSD---YGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRL 500
Db    348 DNPP--EETAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWNAIYRISGGDPAGRF 402

Qy    501 GVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDV 560
Db    403 AIQTDPNSNDGLVTVVKPIDFETNRMYVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDV 462

Qy    561 NEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTG 617
Db    463 NENPYFAPNPKIIRQEEGLHAGTVLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDSVNG 522

Qy    618 EIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTG 673
Db    523 QITTIAVLDRESPNVKANIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVLPQEAE 582

Qy    674 LFFCHPLSAPGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRH 732
Db    583 I--CE-TPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQLNLKI 639

Qy    733 TDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV---GMA 788
Db    640 KFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIVGAGLG 692

Qy    789 VGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
Db    693 TGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 733


Search completed: January 15, 2003, 08:23:09
Job time : 176.909 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:27 ; Search time 61.7352 Seconds
(without alignments)
1295.596 Million cell updates/sec

Title: US-09-079-819-178
Perfect score: 4321
Sequence: 1 MILQAHLHSLCLLMLYLATG.....DKGKDNVESAQASEVKPLRS 832

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4303	99.6	832	2	S55396	LI-cadherin - huma
2	3414	79.0	827	2	A53954	LI-cadherin precur
3	954	22.1	829	2	I46536	Ksp-cadherin - rab
4	715	16.5	913	1	IJCHCR	R-cadherin precurs
5	705	16.3	913	1	A47543	R-cadherin precurs
6	699.5	16.2	906	1	IJHUCN	cadherin 2 precurs
7	696	16.1	912	1	IJCHCN	N-cadherin precurs
8	694.5	16.1	906	1	IJMSCN	N-cadherin precurs
9	688.5	15.9	877	1	IJBOCN	N-cadherin precurs
10	684.5	15.8	916	2	C38992	cadherin 4 precurs
11	677	15.7	783	2	I50116	N-cadherin precurs
12	671.5	15.5	826	2	B55363	desmocollin, type
13	671.5	15.5	896	2	A55363	desmocollin, type
14	667.5	15.4	847	1	IJHUDB	desmocollin 3b pre
15	667.5	15.4	901	1	IJHUDA	desmocollin 3a pre
16	663.5	15.4	906	1	IJXLC2	N-cadherin 2 precu
17	661	15.3	896	2	I45858	desmocollin - bovi
18	657	15.2	713	2	B38992	cadherin 13 precur
19	653.5	15.1	809	1	IJBODD	desmocollin 2b pre
20	653.5	15.1	863	1	IJBODC	desmocollin 2a pre
21	647.5	15.0	905	1	IJXLC1	N-cadherin 1 precu
22	634	14.7	887	1	IJCHCL	E-cadherin precurs
23	630.5	14.6	712	1	IJMSCT	T-cadherin precurs
24	630.5	14.6	717	2	I51206	T-cadherin 2 - chi
25	609	14.1	884	2	S34438	uvomorulin - mouse
26	608	14.1	790	2	G02678	cadherin-14 - huma
27	604	14.0	884	1	IJMSCE	E-cadherin precurs
28	597.5	13.8	732	1	IJCHCB	B-cadherin precurs
29	591	13.7	882	1	IJHUCE	cadherin 1 precurs
30	585.5	13.6	839	1	IJBODF	desmocollin 1b pre
31	580.5	13.4	895	1	IJXLCP	EP-cadherin precur
32	580	13.4	730	1	IJMSCM	M-cadherin - mouse
33	579.5	13.4	871	2	S47518	cadherin - African
34	576	13.3	761	1	IJBODE	desmocollin 1a - b
35	574.5	13.3	796	2	A53584	OB-cadherin precur
36	574.5	13.3	840	2	I37281	Dscla precursor -
37	574.5	13.3	894	2	I37282	Dsclb precursor -
38	572.5	13.2	796	2	I49556	cadherin-11 - mous
39	571.5	13.2	796	2	I48277	cadherin-11 - mous
40	571	13.2	3034	2	T14119	seven-pass transme
41	570	13.2	770	2	B48910	desmocollin 1b pre
42	570	13.2	824	2	A48910	desmocollin 1a pre
43	565	13.1	814	2	G02878	cadherin-15 precur
44	564	13.1	905	2	S43064	cadherin - African
45	560	13.0	822	1	IJMSCP	P-cadherin precurs

ALIGNMENTS

RESULT 1
S55396
LI-cadherin - human
C;Species: Homo sapiens (man)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 31-Mar-2000
C;Accession: S55396
R;Boettinger, A.; Kreft, B.; Fieger, C.; Dlouhy, B.; Berndorff, D.; Goessner, R.; Tau
submitted to the EMBL Data Library, December 1994
A;Description: Molecular cloning of human LI-cadherin:evidence for a novel type of ca
A;Reference number: S55396
A;Accession: S55396
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-832 <BOE>
A;Cross-references: EMBL:X83228; NID:g854174; PIDN:CAA58231.1; PID:g854175
C;Superfamily: cadherin; cadherin repeat homology
F;455-566/Domain: cadherin repeat homology <CR3>

Query Match 99.6%; Score 4303; DB 2; Length 832;
Best Local Similarity 99.5%; Pred. No. 1.3e-262;
Matches 828; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```
Qy       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||||
Db      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIKVKDIN 120

Qy     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
```

```
Db    421 TIEVSDKDFKTLCFVQINVIDINDQTPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGGSSLSSVSEFHLILMDV 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGGSSLSSVSEFHLILMDV 660

Qy    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy    721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
          |||||||||||||:|||| ||||||||||||||||||||||||||||||||||||||||||||
Db    721 INGTHARLSTRHTEFEEREYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832


RESULT 2
A53954
LI-cadherin precursor - rat
N;Alternate names: liver-intestine cadherin
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Oct-1994 #sequence_revision 07-Oct-1994 #text_change 31-Mar-2000
C;Accession: A53954
R;Berndorff, D.; Gessner, R.; Kreft, B.; Schnoy, N.; Lajous-Petter, A.M.; Loch, N.; Reut
J. Cell Biol. 125, 1353-1369, 1994
A;Title: Liver-intestine cadherin: molecular cloning and characterization of a novel Ca(
A;Reference number: A53954; MUID:94266966; PMID:8207063
A;Accession: A53954
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-827 <BER>
A;Cross-references: GB:X78997; NID:g505562; PIDN:CAA55631.1; PID:g505563
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; intestine; liver;
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-827/Product: LI-cadherin #status predicted <MAT>
F;454-565/Domain: cadherin repeat homology <CR3>

  Query Match            79.0%;  Score 3414;  DB 2;  Length 827;
  Best Local Similarity  78.8%;  Pred. No. 1e-206;
  Matches  655;  Conservative   70;  Mismatches  102;  Indels    4;  Gaps    1;

Qy      2 ILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFE 61
          :: | || |||| |||   ||||||||||||||||||:||||||||||||||||||||||
Db      1 MVSAQLHFLCLLTLYLTGAYGQEGKFSGPLKPMTFSIFEGQEPSQIIFQFKANPPAVTFE 60

Qy     62 LTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIND 121
          |||||| || ||::|||||: | |||||||: |:||:|||||: | ||:||||| ||||||||
Db     61 LTGETDGIFKIEKDGLLYHTRVLDRETRAVHHLQLAALDSQGAIVDGPVPIIIEVKDIND 120

Qy    122 NRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQ 181
          |||||||:|||||||||||||||:|||||||||||||||||:|||||||| ||||||||
Db    121 NRPTFLQTKYEGSVRQNSRPGKPFMYVNATDLDDPATPNGQLFYQIVIQLPKINNVMYFQ 180

Qy    182 INNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKA 241
          |:|||||||| |||| |:| ||| ||||:|||||||:||||||||||| | ||||||
Db    181 IDNKTGAISLTPEGSQVLDPIKNPYYNLVVSVKDMGGQNENSFSDTTSVDITVRENIWKA 240

Qy    242 PKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDRE 301
          |:|||: || |||||||||||:|||||| |||::|||||:|||||||||:||||||||||
Db    241 PEPVEIRENLTDPHPIKITQVQWNDPGAHYSLINKEKLPQFPFSIDQEGNIYVTQPLDRE 300

Qy    302 EKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGT 361
          |||::||:| |||| ||||:||||| ||| |||||||| | ||||||||| ||:|||
Db    301 EKDSHVFFATAKDENGKPLAYPLEIRVKVIDINDNPPTCLSQVTVFEVQENEVLGSSIGI 360

Qy    362 LTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLT 421
            ||| || |  || | ||:|:||||:| | |||| | | :|| |:||||||:||||||
Db    361 FAAHDMDEANNINSILKYRLVDQTPKVPSDELFLIDEYGGKVQLGKRSLKKQDSPQYNLT 420

Qy    422 IEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPF 481
          :|||| |||||| :|:|||||||||||||:||||: ||:||| |||||| ||||| ||||
Db    421 VEVSDIDFKTLCSLQVNVIDINDQIPIFERSDYGSKTLSEDTAIGSTILIIQATDDDEPF 480

Qy    482 TGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVF 541
          |||||||| |:::||:|||| | ||| |||||| |:||||||  |::|||||||||||
Db    481 TGSSKILYKIVQGDTEGRLEVVTDPMTNTGYVKIRKPLDFETEPVTSIVFKAENPEPLVN 540

Qy    542 GVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYS 601
          |::|||||||| | | |||||| | | | :||| |||| ||||||| 4||:||||| :|||
Db    541 GIEYNASSFASFELTVTDVNEVPVFPQQIFQANVSEDTAIGTKVGTVTARDPEGLTVSYS 600

Qy    602 LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGGSSLSSVSEFHLILMDVN 661
          |: : ||||||| |||:||| ||||||  | |||||||||||||||| : |||:|||||
Db    601 LKDNKRGWLKIDSVTGDIFSTAPLDRETESVYRVQVVATEVGGGSSLSSTAYFHLVLMDVN 660

Qy    662 DNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKI 721
          |||||||||||||||||||||||||||||||||||||||||  | | |||:||  |||:|||||||
Db    661 DNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQQSVRRPKFTFALGRESLQSDWEVSKI 720

Qy    722 NGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTG 781
          |||||||||:|| |||: | : | ||||||:||:|||||| |||| ||:|||||||:| |
Db    721 NGTHARLSTKHTRFEEQVYDIPILINDGGQPPMEGIVSLSVTFCQCVDGSCFRPAGNQVG 780

Qy    782 IPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
          |||||||||||||| ||||||||||||::|||    ||  |: | |||||
Db    781 IPTVGMAVGILLTTFLVIGIILAVVFIRMRKDK----VEDPQSPENKPLRS 827


RESULT 3
I46536
Ksp-cadherin - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 05-Nov-1999
C;Accession: I46536
R;Thomson, R.B.; Igarashi, P.; Biemesderfer, D.; Kim, R.; Abu-Alfa, A.; Soleimani, M.
J. Biol. Chem. 270, 17594-17601, 1995
A;Title: Isolation and cDNA cloning of Ksp-cadherin, a novel kidney-specific member o
A;Reference number: I46536; MUID:95340560; PMID:7615566
A;Accession: I46536
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-829 <THO>
A;Cross-references: EMBL:U28945; NID:g902885; PIDN:AAC48472.1; PID:g902886

  Query Match            22.1%;  Score 954;  DB 2;  Length 829;
  Best Local Similarity  31.6%;  Pred. No. 5.7e-52;
  Matches  251;  Conservative  142;  Mismatches  354;  Indels   48;  Gaps   22;

Qy     62 LTGET----DNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
          |:|::    :  | :| | | |   |||||| :: : :||     :| :: ||  :|: |
Db     57 LSGDSGVAAEGPFSVEAESGFLLVTRALDREEQAEYQIQVTLEAEDGHVLWGPQSVTVHV 116

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
          || ||  | | |: |   : | :||| |||:: |:| |:| | |  | : |: | |   :
Db    117 KDENDQVPQFSQALYSARLSQGTRPGVPFLFLEASDGDEPGTANSDLRFHILSQTPAQPS 176

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
            |::  : ||::|: |||   : |    | |:: ||||| |: :    | :|:| : |
Db    177 PDVFRLEPRLGALALSPEGSAGFDHALEGPYQLLVQVKDMGDQA-SGHQATATVEISIVE 235

Qy    237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
          : |   :|| : ||   |:|  : || |:     | |   |  | || :| || :|||
Db    236 STWTPLEPVHLAENLKVPYPHHLAQVHWSGGDVHYRL---ESQPPGPFDVDTEGKLYVTG 292
```

```
Qy    297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCP---SPVTVFEVQENE 353
          |||| :: ||    |::  |:  : |||:|| | | ||: | ||    ||:| |:
Db    293 ELDREAQEQYVLQVQAQNSRGEDYAEPLELHVVVTDENDHAPVCPPRGPPVSVPELSPP- 351

Qy    354 RLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKL-PMDGLFLIQTYAGMLQLAKQSLKK 412
            | :: ||:| | |   : || : ||::   |:  |  | | :   :| : |    |:
Db    352 --GTAVTTLSAEDADAPGSPNSHVVYRLLSPEPQEGPEGGAFQLDPTSGSVSLGAAPLEA 409

Qy    413 QDTPQYNLTIEV-------SDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNI 465
              |: ::|       ::    : | | : | |:||  | |  |  | ::| |||
Db    410 GQ----NMLLQVLAVDLAGAEAGLSSTCEVAVTVTDVNDHAPEFTSSQVGPVSLPEDTEP 465

Qy    466 GSTILTIQATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFET 523
          |: : |: ||||| ||   : ::: : |  || ||  |:| :| :    : : | | :|
Db    466 GTLVATLTATDADLEP---AFRLMDFTIEAGDGEGTFGLDWEPDSGHVQLYLLKNLSYEA 522

Qy    524 AAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGT 583
          |    :|    |    | |      : |  |::|  |    |:  |  ::| :  :   |:
Db    523 APSHTVVVVVRNVVETV-GPGPGPGATATVTVLVEKVMPPPRLEQKSYEADIPVNAPAGS 581

Qy    584 KVGNVTAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEV 642
            :  :   :|    : :|| |: ||  |  |:||: :  ||     |  | | | | :
Db    582 FLLTIQPAEPWNGALRFSLVNDSEGWFCIQKVSGEVHTARPLQGARPGDSYTVLVEAQDA 641

Qy    643 GGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDD-DQHLFRGP 701
              ||: :   :  :          |          | |    | ||    ::| |     ||
Db    642 DAPRLSTSAALVIHFLRAPPAPALPLAPMPSRHLCTPRQDHGVLIPAPSEDPDMATGHGP 701

Qy    702 HFTFSLG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSL 760
          ::|:|| : ::| || :  :| :|| |:      | | ::| : ::   |     :  |
Db    702 -YSFALGPNPTVQRDWRLQPLNDSHAFLTLALHWVEPREHIVPVVVSQDAR-----VWQL 755

Qy    761 P--VTFCSC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKD 817
          |  |  | |  || | |  |   |:||   |||||: ||  ||  | ::|  :   : ||
Db    756 PVRVVVCRCNTEGECMRKVGRMKGMPTKLSAVGILVGTLAAIGFFLILIFTHLALARKKD 815

Qy    818 NVESAQASEVKPLRS 832
              | |  | ||::
Db    816 --LDAPADNV-PLKA 827


RESULT 4
IJCHCR
R-cadherin precursor - chicken
C;Species: Gallus gallus (chicken)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 16-Jun-2000
C;Accession: JH0424
R;Inuzuka, H.; Miyatani, S.; Takeichi, M.
Neuron 7, 69-79, 1991
A;Title: R-cadherin: a novel Ca2+-dependent cell-cell adhesion molecule expressed in the
A;Reference number: JH0424; MUID:91299341; PMID:1712604
A;Accession: JH0424
A;Molecule type: mRNA
A;Residues: 1-913 <INU>
A;Cross-references: GB:D14459; GB:D00849; NID:g222854; PIDN:BAA03356.1; PID:g222855
A;Experimental source: retina
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought t
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; retina; transmemb
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-166/Domain: propeptide #status predicted <PRO>
F;167-913/Product: R-cadherin #status predicted <MAT>
F;167-721/Domain: extracellular #status predicted <EXT>
F;169-274/Domain: cadherin repeat homology <CR1>
F;244-249/Region: cadherin binding #status predicted
F;277-389/Domain: cadherin repeat homology <CR2>
F;392-504/Domain: cadherin repeat homology <CR3>
F;507-612/Domain: cadherin repeat homology <CR4>
F;613-721/Domain: cadherin repeat homology <CR5>
F;722-753/Domain: transmembrane #status predicted <TMM>
F;754-913/Domain: intracellular #status predicted <INT>
F;870-885/Region: serine-rich
F;280,409,554,629,658,699/Binding site: carbohydrate (Asn) (covalent) #status predicted


  Query Match           16.5%;  Score 715;  DB 1;  Length 913;
  Best Local Similarity  26.9%;  Pred. No. 7.2e-37;
  Matches  203;  Conservative  117;  Mismatches  329;  Indels  106;  Gaps   19;

Qy    123 RPTFLQSKYEGSVRQNSRPGKPFLYV------------NATDLDDPATPNGQLYYQIVI 169
          :| | :  |   | ||   |:  | |                ||     :| :|   |
Db     31 KPGFSEEDYTAFVSQNIMEGQKLLKVKFNNCAGNKGVRYETNSLDFKVRADGTMY--AVH 88

Qy    170 QLPMINNVMYFQI------------------------------NNKTGAISLTREGSQE 198
          |: | :  :   :                                 |:| : | :: |
Db     89 QVQMASKQLILMVTAWDPQTLGRWEAIVRFLVGEKLQHNGHKPKGRKSGPVDLAQQQSDT 148

Qy    199 LNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIK 258
          | | :          | | :  |  :                 |  | |: : |||  | | :
Db    149 LLPWRQHQ-----SAKGLRRQKRD---------------WVIP-PINVPENSRGPFPQQ 186

Qy    259 ITQVRWN---DPGAQYSL--VDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVA 312
          : ::| :   :   :||:  |  :: |   ||||   | :|||:|:||||: :|   | |
Db    187 LVRIRSDKDKEIHIRYSITGVGADQPPMEVFSIDPVSGRMYVTRPMDREERASYHLRAHA 246

Qy    313 KDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENT 372
          |  |  :  |:::::| | |:||| |   : |    | |  : |  : |:||:| |:  |
Db    247 VDMNGNKVENPIDLYIYVIDMNDNRPEFINQVYNGSVDEGSKPGTYVMTVTANDADDSTT 306

Qy    373 ANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD---- 428
          ||  : |||| |||: |   :| | :  | :       | ::   || : :: :| :
Db    307 ANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGLDREKVQQYMVIVQATDMEGNLN 366

Qy    429 --FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
                 | | |:||  | |  | |       :  :   :  :   | |:| : :
Db    367 YGLSNTATAIITVTDVNDNPPEFTTSTYSG--EVPENRVEVVVANLTVMDRDQPHSPNWN 424

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
          :| || ||  |   : ||| || | | : |:|:|      :      |  ||  |:: :
Db    425 AIYRIISGDPSGHFTIRTDPVTNEGMVTVVKAVDYEMNRAFMLTVMVSNQAPLASGIQMS 484

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLR 603
            | |  |: ||||||||| |  :    :: | |  || :   :| ||:      : ||
Db    485 FQSTAGVTISVTDVNEAPYFPTNHKLIRLEEGVPTGTVLTTFSAVDPDRFMQQAVRYSKL 544

Qy    604 GDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMD 659
          |   || |:   |:| : | ||||:     : |    :| : |    |      : |:|
Db    545 SDPANWLNINATNGQITTAAVLDRESDYIKNNVYEATFLAADNGIPPASGTGTLQIYLID 604

Qy    660 VNDNPPRLAKDYTGLFFCHPLSAPGSLIFEAT----DDDQHLFRGPHFTFSLGS--GSLQ 713
          :||| | |      :       |   :   |   | | ::  || | | | |   :::
Db    605 INDNAPELLPKEAQI-----CEKPNLNVINITAADADIDPNV--GP-FVFELPSVPSAVR 656

Qy    714 NDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCF 773
          :| ::::|| :|:|| |    |   | | | : | | |||     : |  | | |
Db    657 KNWTITRLNGDYAQLSLRIMYLEAGVYDVPIIVTDSGNPPLYNTSIIKVKVCPCDENGDC 716

Qy    774 RPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFI 808
            |    :   |:  | ::  |: | |:| :| :
Db    717 TTIG---AVAAAGLGTGAIIAILICIIILLTMVLL 748


RESULT 5
A47543
R-cadherin precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-May-1994 #sequence_revision 26-May-1994 #text_change 22-Jun-1999
C;Accession: A47543; I55501
R;Hutton, J.C.; Christofori, G.; Chi, W.Y.; Edman, U.; Guest, P.C.; Hanahan, D.; Kell
Mol. Endocrinol. 7, 1151-1160, 1993
A;Title: Molecular cloning of mouse pancreatic islet R-cadherin: differential express
A;Reference number: A47543; MUID:94067164; PMID:8247017
A;Accession: A47543
A;Molecule type: mRNA
A;Residues: 1-913 <HUT>
A;Cross-references: EMBL:X69966; NID:g429111; PIDN:CAA49589.1; PID:g429112
R;Matsunami, H.; Miyatani, S.; Inoue, T.; Copeland, N.; Gilbert, D.; Jenkins, N.; Tak
```

```
J. Cell Sci. 106, 401-409, 1993
A;Title: Cell binding specificity of mouse R-cadherin and chromosomal mapping of the gen
A;Reference number: I55501; MUID:94095672; PMID:8270638
A;Accession: I55501
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-913 <RES>
A;Cross-references: GB:D14888; NID:g457658; PIDN:BAA03605.1; PID:g457659
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-166/Domain: propeptide #status predicted <PRO>
F;167-913/Product: R-cadherin #status predicted <MAT>
F;167-721/Domain: extracellular #status predicted <EXT>
F;169-274/Domain: cadherin repeat homology <CR1>
F;244-249/Region: cadherin binding #status predicted
F;277-389/Domain: cadherin repeat homology <CR2>
F;300-304/Domain: calcium binding #status predicted <CAB>
F;392-504/Domain: cadherin repeat homology <CR3>
F;507-612/Domain: cadherin repeat homology <CR4>
F;613-721/Domain: cadherin repeat homology <CR5>
F;722-753/Domain: transmembrane #status predicted <TMM>
F;754-913/Domain: intracellular #status predicted <INT>
F;870-885/Region: serine-rich
F;280,409,554,629,658,699/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match            16.3%;  Score 705;  DB 1;  Length 913;
  Best Local Similarity  30.1%;  Pred. No. 3.1e-36;
  Matches  178;  Conservative  101;  Mismatches  280;  Indels   33;  Gaps   12;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
Db    168 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFNIDSMSGRM 226

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
Db    227 YVTRPMDREERASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNRPEFINQVYNGSVDEG 286

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
Db    287 SKPGTYVMTVTANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGLDR 346

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
Db    347 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPPEFTTSTFAG--EVPENRIE 404

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
Db    405 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 464

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
Db    465 FMLTVMVSNQAPLASGIQMSFQSTAGVTISVTDVNEAPYFPSNHKLIRLEEGVPAGTALT 524

Qy    587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
Db    525 TFSAVDPDRFMQQAVRYSKLSDPANWLHINTSNGQITTAAILDRESLYTKNNVYEATFLA 584

Qy    640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPG--SLIFEATDDDQHL 697
Db    585 ADNGIPPASGTGTLQIYLIDINDNAPQLLPKEAQI-----CERPGLNAINITAADADMDP 639

Qy    698 FRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEG 756
Db    640 NIGPYVFELPFIPTTVRKNWTITRLNGDYAQLSLRILYLEAGVYDVPIIVTDSGNPPLSN 699

Qy    757 IVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFI 808
Db    700 TSVIKVKVCPCDENGDCTTVG---AVAAAGLGTGAIVAILICIVILLIMVLL 748


RESULT 6
```

```
IJHUCN
cadherin 2 precursor - human
N;Alternate names: N-cadherin; neuronal cadherin
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A38870; S11487; JQ0751; S13799
R;Reid, R.A.
submitted to the EMBL Data Library, November 1990
A;Reference number: A38870
A;Accession: A38870
A;Molecule type: mRNA
A;Residues: 1-906 <REI>
A;Cross-references: EMBL:X54315; NID:g34998; PIDN:CAA38213.1; PID:g34999
R;Reid, R.A.; Hemperly, J.J.
Nucleic Acids Res. 18, 5896, 1990
A;Title: Human N-cadherin: nucleotide and deduced amino acid sequence.
A;Reference number: S11487; MUID:91016946; PMID:2216790
A;Accession: S11487
A;Molecule type: mRNA
A;Residues: 1-340,'N',342-698,'R',700-704,'F',706-906 <RE2>
A;Cross-references: EMBL:X54315
A;Note: this sequence has been revised in reference A38870
R;Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell, D.; Spurr, N.; Goodfello
J. Neurochem. 55, 805-812, 1990
A;Title: N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadhe
A;Reference number: JQ0751; MUID:90347462; PMID:2384753
A;Accession: JQ0751
A;Molecule type: mRNA
A;Residues: 160-194,'IR',197-211,'L',213-227,'Q',229,'N',231-235,'G',237-248,'T',250-
A;Cross-references: GB:M34064
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought
C;Genetics:
A;Gene: GDB:CDH2; NCAD
A;Cross-references: GDB:128185; OMIM:114020
A;Map position: 18q12.1-18q12.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-159/Domain: propeptide #status predicted <PRO>
F;160-906/Product: N-cadherin #status predicted <MAT>
F;160-714/Domain: extracellular #status predicted <EXT>
F;162-267/Domain: cadherin repeat homology <CR1>
F;237-242/Region: cadherin binding #status predicted
F;270-382/Domain: cadherin repeat homology <CR2>
F;385-497/Domain: cadherin repeat homology <CR3>
F;500-605/Domain: cadherin repeat homology <CR4>
F;606-712/Domain: cadherin repeat homology <CR5>
F;715-746/Domain: transmembrane #status predicted <TMM>
F;747-906/Domain: intracellular #status predicted <INT>
F;865-878/Region: serine-rich
F;190,273,325,402,572,622,651,692/Binding site: carbohydrate (Asn) (covalent) #status

  Query Match            16.2%;  Score 699.5;  DB 1;  Length 906;
  Best Local Similarity  27.4%;  Pred. No. 6.7e-36;
  Matches  210;  Conservative  122;  Mismatches  331;  Indels  103;  Gaps   21;

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
Db     47 KDVHEGQP-LLNVKF-----SNCNGKRKVQYESSEPADFKVDEDGMVY--AVRSFPLSSE 98

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
Db     99 HAKFLIY-----------AQDKETQEKWQVAVKLSLKPT--LTEESVKESAEVEEIVFP 144

Qy    235 ------------TENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQ 270
Db    145 RQFSKHSGHLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYTVTGPGA- 202

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
Db    203 ------DQPPTGIFIINPISGQLSVTKPLDREQIARFHLRAHAVDINGNQVENPIDIVIN 256

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
```

```
Db    257 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443

Db    317 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGIPTYGLSNTATAVITVTDVN 376

Qy    444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502

Db    377 DNPPEFTAMTFYGEV---PENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAI 433

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562

Db    434 QTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 493

Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619

Db    494 NPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQI 553

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675

Db    554 TTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL------- 606

Qy    676 FCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727

Db    607 ---PQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQ 663

Qy    728 LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV- 785

Db    664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIV 716

Qy    786 --GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829

Db    717 GAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762


RESULT 7
IJCHCN
N-cadherin precursor, neuronal - chicken
C;Species: Gallus gallus (chicken)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A29964
R;Hatta, K.; Nose, A.; Nagafuchi, A.; Takeichi, M.
J. Cell Biol. 106, 873-881, 1988
A;Title: Cloning and expression of cDNA encoding a neural calcium-dependent cell adhesic
A;Reference number: A29964; MUID:88153917; PMID:2831236
A;Accession: A29964
A;Molecule type: mRNA
A;Residues: 1-912 <HAT>
A;Cross-references: GB:X07277; NID:g63649; PIDN:CAA30258.1; PID:g63650
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought t
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane prc
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-164/Domain: propeptide #status predicted <PRO>
F;165-912/Product: N-cadherin #status predicted <MAT>
F;165-720/Domain: extracellular #status predicted <EXT>
F;167-272/Domain: cadherin repeat homology <CR1>
F;242-247/Region: cadherin binding #status predicted
F;275-387/Domain: cadherin repeat homology <CR2>
F;390-502/Domain: cadherin repeat homology <CR3>
F;505-611/Domain: cadherin repeat homology <CR4>
F;612-720/Domain: cadherin repeat homology <CR5>
F;721-752/Domain: transmembrane #status predicted <TMM>
F;753-912/Domain: intracellular #status predicted <INT>
F;869-884/Region: serine-rich
F;278,330,407,578,628,657/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match           16.1%;  Score 696;  DB 1;  Length 912;
  Best Local Similarity  26.9%;  Pred. No. 1.1e-35;
  Matches  209;  Conservative  117;  Mismatches  328;  Indels  124;  Gaps   22;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183

Db     41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy    184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENSFSDTT 228

Db     82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy    229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269

Db    137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy    270 QYSLVDKEKLPRFPFSI-----DQ----------EGDIYVTQPLDREEKDAYVFYAVAK 313

Db    190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy    314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373

Db    246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy    374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428

Db    306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy    429 -FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486

Db    366 GLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWN 422

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546

Db    423 ARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISYSL 602

Db    483 PQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRYSK 542

Qy    603 RGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILM 658

Db    543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIYLL 602

Qy    659 DVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQND 715

Db    603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657

Qy    716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFR 774

Db    658 WTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC-- 715

Qy    775 PAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829

Db    716 -----TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768


RESULT 8
IJMSCN
N-cadherin precursor, neuronal - mouse
C;Species: Mus musculus (house mouse)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: A32759; A46163
R;Miyatani, S.; Shimamura, K.; Hatta, M.; Nagafuchi, A.; Nose, A.; Matsunaga, M.; Hat
Science 245, 631-635, 1989
A;Title: Neural cadherin: role in selective cell-cell adhesion.
A;Reference number: A32759; MUID:89346748; PMID:2762814
A;Accession: A32759
A;Molecule type: mRNA
A;Residues: 1-906 <MIY>
A;Cross-references: GB:M31131; NID:g192327; PIDN:AAA37353.1; PID:g309125
R;Miyatani, S.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Takeichi, M.
Proc. Natl. Acad. Sci. U.S.A. 89, 8443-8447, 1992
A;Title: Genomic structure and chromosomal mapping of the mouse N-cadherin gene.
A;Reference number: A46163; MUID:92409532; PMID:1528849
A;Accession: A46163
A;Status: preliminary
A;Molecule type: DNA
```

```
A;Residues: 839-906 <MI2>
A;Cross-references: GB:S45011; NID:g256010; PIDN:AAB23356.1; PID:g256011
A;Note: sequence extracted from NCBI backbone (NCBIN:113759, NCBIP:113760)
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought t
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-159/Domain: propeptide #status predicted <PRO>
F;160-906/Product: N-cadherin #status predicted <MAT>
F;160-714/Domain: extracellular #status predicted <EXT>
F;162-267/Domain: cadherin repeat homology <CR1>
F;237-242/Region: cadherin binding #status predicted
F;270-382/Domain: cadherin repeat homology <CR2>
F;385-497/Domain: cadherin repeat homology <CR3>
F;500-605/Domain: cadherin repeat homology <CR4>
F;606-714/Domain: cadherin repeat homology <CR5>
F;715-746/Domain: transmembrane #status predicted <TMM>
F;747-906/Domain: intracellular #status predicted <INT>
F;865-878/Region: serine-rich
F;190,273,325,402,572,651,692/Binding site: carbohydrate (Asn) (covalent) #status predic

  Query Match             16.1%;  Score 694.5;  DB 1;  Length 906;
  Best Local Similarity   27.6%;  Pred. No. 1.4e-35;
  Matches  211;  Conservative  122;  Mismatches  332;  Indels   99;  Gaps   23;

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
          ||::: :|  |  |:    |:     :   | ::   |     :| :|   |   |:
Db     47 KDVHEGQP-LLNVKFSNCNRK-----RKVQYESSEPADFKVDEDGTVY--AVRSFPLTAE 98

Qy    177 VMYFQI---NNKTG-----AISLTREGSQELNPAKNP-SYNLVISVKDM---GGQSENSF 224
             | |   : :|      |::|:|| :    | | |     ::  : :    |  :
Db     99 QAKFLIYAQDKETQEKWQVAVNLSREPTLTEEPMKEPHEIEEIVFPRQLAKHSGALQRQK 158

Qy    225 SDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQYS 272
          |           |  | |: : |||  | | :: ::| :             |||
Db    159 RD-----------WVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA--- 202

Qy    273 LVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVK 331
             :: |   | |:   | : ||:|||||    :   | | |  |  :  |::| : |
Db    203 ----DQPPTGIFIINPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVINVI 258

Qy    332 DINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMD 391
          |:||| |      |    | |  : |  : |:|| | |: |  |  | |||: | |  |
Db    259 DMNDNRPEFLHQVWNGSVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTPSP 318

Qy    392 GLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQ 445
           :| |    | :     | ::   || | |: :| :              | | |:||
Db    319 NMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVNDN 378

Qy    446 IPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDT 504
          | |   : || :    :  :   :  :  || |:| | :    | |  ||  ||  : |
Db    379 PPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWNAAYRISGGDPTGRFAILT 435

Qy    505 DPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAP 564
          ||:: | | : ||:||||  :  :   |||  ||  |:::   | |  :: | |||| |
Db    436 DPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNENP 495

Qy    565 QFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFS 621
           |: :   :  | :  || :  :||:||:     :| |:   |  ||||| | |:| :
Db    496 YFAPNPKIIRQEEGLHAGTMLTTLTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQITT 555

Qy    622 VAPLDREA----GSPYRVQVVATEVGGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFC 677
          :| ||||:     : |    :|:: |   :|      : |:|:||| |::
Db    556 IAVLDRESPYVQNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL--------- 606

Qy    678 HPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLS 729
           |  |        | |:   | | |      ||  |   |   ::: :| :::::||  |:|:
Db    607 -PQEAETCETPEPNSINIAALDYDIDPNAGPFAFDLPLSPVTIKRNWTINRLNGDFAQLN 665

Qy    730 TRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV--- 785
           :   |  | | | | | | ||   |  | |  | |   | |       | :  :
Db    666 LKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIVGA 718

Qy    786 GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
          |:  | ::  || | |:| :| : :   | :|    |:   : |
Db    719 GLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762


RESULT 9
IJBOCN
N-cadherin precursor - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C;Accession: S11693
R;Liaw, C.W.; Cannon, C.; Power, M.D.; Kiboneka, P.K.; Rubin, L.L.
EMBO J. 9, 2701-2708, 1990
A;Title: Identification and cloning of two species of cadherins in bovine endothelial
A;Reference number: S11693; MUID:90360979; PMID:2390969
A;Accession: S11693
A;Molecule type: mRNA
A;Residues: 1-877 <LIA>
A;Cross-references: EMBL:X53615; NID:g164; PIDN:CAA37677.1; PID:g664894
C;Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are though
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F;1-130/Domain: propeptide (fragment) #status predicted <PRO>
F;131-877/Product: N-cadherin #status predicted <MAT>
F;131-685/Domain: extracellular #status predicted <EXT>
F;133-238/Domain: cadherin repeat homology <CR1>
F;208-213/Region: cadherin binding #status predicted
F;241-353/Domain: cadherin repeat homology <CR2>
F;356-468/Domain: cadherin repeat homology <CR3>
F;471-576/Domain: cadherin repeat homology <CR4>
F;577-685/Domain: cadherin repeat homology <CR5>
F;686-717/Domain: transmembrane #status predicted <TMM>
F;718-877/Domain: intracellular #status predicted <INT>
F;836-849/Region: serine-rich
F;161,244,296,373,543,593,622,663/Binding site: carbohydrate (Asn) (covalent) #status

  Query Match             15.9%;  Score 688.5;  DB 1;  Length 877;
  Best Local Similarity   26.7%;  Pred. No. 3.1e-35;
  Matches  203;  Conservative  126;  Mismatches  343;  Indels   87;  Gaps   19;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLP---------MINN 176
          | :  |   : ::   |:| | |  ::       ||:   |     |         |:
Db      7 FPEDVYSAVLSRDVLEGQPLLNVKFSNC------NGKRKVQYESSEPADFKVDEDGMVYA 60

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
          |  | ::::     :  :  :          : : : :          |:|  ::  :: ||
Db     61 VRSFPLSSEHSKFLIYAQDKE-----TQEKWQVAVKLSLKPALPEDSVKESREIEEIVFP 115

Qy    235 ------------TENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQ 270
                      :  |  | |: : |||  | | :: ::| :             |||
Db    116 RQVTKHNGYLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA- 173

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
                ::  |   | |:   | : ||:|||||    :   | | |  |  :  |::| :
Db    174 ------DQPPTGIFIINPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVIN 227

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
          | |:||| |      |    | |  : |  : |:|| | |: |  |  | |||: | |  |
Db    228 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTP 287

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
             :| |     | :     | ::   || | |: :| :              | | |:|
Db    288 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVN 347

Qy    444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502
          |  | |   : || :    :  :   :  :  || |:| | :   :| |  ||  ||  :
Db    348 DNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWNAIYRISGGDPAGRFAI 404

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
           |||::| | | : ||:||||  :  :   |||  ||  |:::   | |  :: | ||||
Db    405 QTDPNSNDGLVTVVKPIDFETNRMYVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 464

Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
```

```
Db    465 NPYFAPNPKIIRQEEGLHAGTVLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDSVNGQI 524

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675

Db    525 TTIAVLDRESPNVKANIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVLPQEAEI- 583

Qy    676 FCHPLSAPGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 734

Db    584 -CE-TPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQLNLKIKF 641

Qy    735 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV---GMAVG 790

Db    642 LEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIVGAGLGTG 694

Qy    791 ILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829

Db    695 AIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 733


RESULT 10
C38992
cadherin 4 precursor - human
N;Alternate names: R-cadherin
C;Species: Homo sapiens (man)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 20-Aug-1999
C;Accession: C38992
R;Suzuki, S.; Sano, K.; Tanihara, H.
Cell Regul. 2, 261-270, 1991
A;Title: Diversity of the cadherin family: evidence for eight new cadherins in nervous t
A;Reference number: S24305; MUID:91283540; PMID:2059658
A;Accession: C38992
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-916 <SUZ>
A;Cross-references: GB:L34059; NID:g506409; PIDN:AAA35627.1; PID:g506410
C;Genetics:
A;Gene: GDB:CDH4
A;Cross-references: GDB:622850
A;Map position: 16q24.1-16qter
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-166/Domain: propeptide #status predicted <PRO>
F;167-913/Product: R-cadherin #status predicted <MAT>
F;167-721/Domain: extracellular #status predicted <EXT>
F;169-274/Domain: cadherin repeat homology <CR1>
F;244-249/Region: cadherin binding #status predicted
F;277-389/Domain: cadherin repeat homology <CR2>
F;300-304/Domain: calcium binding #status predicted <CAB>
F;392-504/Domain: cadherin repeat homology <CR3>
F;507-612/Domain: cadherin repeat homology <CR4>
F;613-721/Domain: cadherin repeat homology <CR5>
F;722-753/Domain: transmembrane #status predicted <TMM>
F;754-913/Domain: intracellular #status predicted <INT>
F;870-885/Region: serine-rich
F;283,412,557,632,661,702/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             15.8%;  Score 684.5;  DB 2;  Length 916;
  Best Local Similarity   29.3%;  Pred. No. 6e-35;
  Matches  176;  Conservative  106;  Mismatches  288;  Indels   31;  Gaps   12;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292

Db    171 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFSINSMSGRM 229

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352

Db    230 YVTRPMDREEHASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNHPEFINQVYNCSVDEG 289

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412

Db    290 SKPGTYVMTITANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466

Db    350 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPSEFTASTFAG--EVPENSVE 407

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526

Db    408 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 467

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586

Db    468 FMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPYFPSNHKLIRLEEGVPPGTVLT 527

Qy    587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639

Db    528 TFSAVDPDRFMQQAVRYSKLSDPASWLHINATNGQITTVAVLDRESLYTKNNVYEATFLA 587

Qy    640 TEVGGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFR 699

Db    588 ADNGIPPASGTGTLQIYLIDINDNAPELLPKEAQICERPNLNA---INITAADADVHPNI 644

Qy    700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIV 758

Db    645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPIIVTDSGNPPLSNTS 704

Qy    759 SLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--FIRIKKDKGK 816

Db    705 IIKVKVCPCDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTMVLLFVMWMKRREK 761

Qy    817 D 817

Db    762 E 762


RESULT 11
I50116
N-cadherin precursor - zebra fish
C;Species: Brachydanio rerio (zebra fish)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jan-2000
C;Accession: I50116
R;Bitzur, S.; Kam, Z.; Geiger, B.
Dev. Dyn. 201, 121-136, 1994
A;Title: Structure and distribution of N-cadherin in developing zebrafish embryos: mo
A;Reference number: I50116; MUID:95178741; PMID:7873785
A;Accession: I50116
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-783 <BIT>
A;Cross-references: EMBL:X67648; NID:g623209; PIDN:CAA47890.1; PID:g623210
C;Superfamily: cadherin; cadherin repeat homology
F;146-258/Domain: cadherin repeat homology <CDH>

  Query Match             15.7%;  Score 677;  DB 2;  Length 783;
  Best Local Similarity   30.1%;  Pred. No. 1.4e-34;
  Matches  187;  Conservative   98;  Mismatches  288;  Indels   48;  Gaps   15;

Qy    239 WKAPKPVEMVENSTDPHP---IKITQVRWNDPGAQYSLV--DKEKLPRFPFSIDQ-EGDI 292

Db     37 WVIP-PVNVLENSRKQFPEELVKIQSDKDKSNTLRYSVTGPGADQNPTGLFIIDPISGLL 95

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352

Db     96 SVTKPLDREHIPNFHLRAHAVDINGNQMENPIDIIINVIDMNDNRPEFTHQIWNGTVDEG 155

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412

Db    156 AKPGTFVMTVTSQDKDDPNTANGMLRYKILSQTPESPSSNMFTINNKTGKIITVAAGLDR 215

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466

Db    216 EKVPQYTLIIQATDMEGNPTYGLSNTATAVIRLLDVNDNAPEFTRETFHGEVPENRVNVI 275

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
```

```
Db    276 VTNLTV--TDKDEPGTPAWNAVYRIISGDPTGRFSIPTDPVTNEGLVTVVKPVDFEMNRS 333

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586

Db    334 FMLTVVADNEVPLASGIHRTRQSTATVSIRVIDVNESPNFDPNPKQIKLEEGLPQWSMLT 393

Qy    587 NVTA-KDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVV 638

Db    394 TFTAHDDPDRYMQQTISYSKLYDPANLLEIDPNNGRISTIAVLDRESPYVKNNLYNATFM 453

Qy    639 ATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSA-------PGSLIFEAT 691

Db    454 ASDNGVPRASGTGTLQIYLLDINDNAPRV----------FPQEAEVCERPEPNAINITAV 503

Qy    692 DDDQHLFRGPHFTFSLGS--GSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDG 749

Db    504 DGDLNPNAGP-YAFELPNRPSDIRRNWTLTRISGDHAQLSLKISYLESGIYELPISITDS 562

Qy    750 GRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFI 808

Db    563 GNLPMSNTTYLRIKVCQCDHHGDCV----DMERIMAAGLGTGAIIAILICIIILLVLVLM 618

Qy    809 RIKKDKGKDNVESAQASEVKP 829

Db    619 FVMWMKRRDKERQAKQLLIDP 639


RESULT 12
B55363
desmocollin, type 4, short form precursor - human
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 23-May-1997
C;Accession: B55363
R;Kawamura, K.; Watanabe, K.; Suzuki, T.; Yamakawa, T.; Kamiyama, T.; Nakagawa, H.; Tsur
J. Biol. Chem. 269, 26295-26302, 1994
A;Title: cDNA cloning and expression of a novel human desmocollin.
A;Reference number: A55363; MUID:95014464; PMID:7929347
A;Accession: B55363
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-826 <KAW>
A;Cross-references: GB:D17427
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; transmembrane protein
F;138-243/Domain: cadherin repeat homology <CR1>
F;246-355/Domain: cadherin repeat homology <CR2>

  Query Match             15.5%;  Score 671.5;  DB 2;  Length 826;
  Best Local Similarity   30.8%;  Pred. No. 3.3e-34;
  Matches  187;  Conservative  106;  Mismatches  263;  Indels   51;  Gaps   22;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQEG 290

Db    137 W-APIPCSMQENSLGPFPLFLQQVE-SDAAQNYTVFYSISGRGVDKEPLNLFYIERD-TG 193

Qy    291 DIYVTQPLDREEKDAY--VFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFE 348

Db    194 NLFCTRPVDREEYDVFDLIAYASTADGYSADL--PLPLPIRVEDENDNHPVFTEAIYNFE 251

Qy    349 VQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQ 408

Db    252 VLESSRPGTTVGVVCATDRDEPDTMHTRLKYSILQQTPRSP--GLFSVHPSTGVITTVSH 309

Qy    409 SLKKQDTPQYNLTIEVSDKDFK-------TLCFVQINVIDINDQIPIFEKSDYGNLTLAE 461

Db    310 YLDREVVDKYSLIMKVQDMDGQFFGLIGTSTCI--ITVTDSNDNAPTFRQNAYE--AFVE 365

Qy    462 DTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDF 521

Db    366 ENAFNVEILRIPIEDKDLINTANWRVNFTILKGNENGHFKISTDKETNEGVLSVVKPLNY 425

Qy    522 ETAAVSNIVFKAENPEPLVFGV-KYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVA 580

Db    426 EENRQVNLEIGVNNEAPFARDIPRVTALNRALVTVHVRDLDEGPECTPAAQYVRIKENLA 485

Qy    581 IGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREAGSP----Y 633

Db    486 VGSKINGYKAYDPENRNGNGLRYKKLHDPKGWITIDEISGSIITSKILDREVETPKNELY 545

Qy    634 RVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDD 693

Db    546 NITVLAIDKDDRSCTGTLAVN--IEDVNDNPPEILQEY--VVICKPKMGYTDIL--AVDP 599

Qy    694 DQHLFRGPHFTFSLGSGS--LQNDWEVSKINGTHARLS-TRHTDFEERAYVVLIRINDGG 750

Db    600 DEPVHGAP-FYFSLPNTSPEISRLWSLTKVNDTAARLSYQKNAGFQE--YTIPITVKD-- 654

Qy    751 RPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAV-GILLTTLLVIGIILAVVFIR 809

Db    655 RAGQAATKLLRVNLCECTHPTQCRATSRSTGVILGKWAILAILLGIALLFSVLLTLVCGV 714

Qy    810 IKKDKGK 816

Db    715 FGATKGK 721


RESULT 13
A55363
desmocollin, type 4, long form precursor - human
C;Species: Homo sapiens (man)
C;Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 20-Jun-2000
C;Accession: A55363
R;Kawamura, K.; Watanabe, K.; Suzuki, T.; Yamakawa, T.; Kamiyama, T.; Nakagawa, H.; T
J. Biol. Chem. 269, 26295-26302, 1994
A;Title: cDNA cloning and expression of a novel human desmocollin.
A;Reference number: A55363; MUID:95014464; PMID:7929347
A;Accession: A55363
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-896 <KAW>
A;Cross-references: GB:D17427; NID:g639672; PIDN:BAA04249.1; PID:g685219
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; transmembrane protein
F;138-243/Domain: cadherin repeat homology <CR1>
F;246-355/Domain: cadherin repeat homology <CR2>

  Query Match             15.5%;  Score 671.5;  DB 2;  Length 896;
  Best Local Similarity   30.8%;  Pred. No. 3.8e-34;
  Matches  187;  Conservative  106;  Mismatches  263;  Indels   51;  Gaps   22;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQEG 290

Db    137 W-APIPCSMQENSLGPFPLFLQQVE-SDAAQNYTVFYSISGRGVDKEPLNLFYIERD-TG 193

Qy    291 DIYVTQPLDREEKDAY--VFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFE 348

Db    194 NLFCTRPVDREEYDVFDLIAYASTADGYSADL--PLPLPIRVEDENDNHPVFTEAIYNFE 251

Qy    349 VQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQ 408

Db    252 VLESSRPGTTVGVVCATDRDEPDTMHTRLKYSILQQTPRSP--GLFSVHPSTGVITTVSH 309

Qy    409 SLKKQDTPQYNLTIEVSDKDFK-------TLCFVQINVIDINDQIPIFEKSDYGNLTLAE 461

Db    310 YLDREVVDKYSLIMKVQDMDGQFFGLIGTSTCI--ITVTDSNDNAPTFRQNAYE--AFVE 365

Qy    462 DTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDF 521

Db    366 ENAFNVEILRIPIEDKDLINTANWRVNFTILKGNENGHFKISTDKETNEGVLSVVKPLNY 425

Qy    522 ETAAVSNIVFKAENPEPLVFGV-KYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVA 580

Db    426 EENRQVNLEIGVNNEAPFARDIPRVTALNRALVTVHVRDLDEGPECTPAAQYVRIKENLA 485

Qy    581 IGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREAGSP----Y 633
```

```
          :|:|:    | |||   |  : |    | :||: || ::| | :   ||||  :|    |
Db    486 VGSKINGYKAYDPENRNGNGLRYKKLHDPKGWITIDEISGSIITSKILDREVETPKNELY 545

Qy    634 RVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDD 693
          : |:| :    | :     :  : ||||||| : ::|  :  | |      ::  | |
Db    546 NITVLAIDKDDRSCTGTLAVN--IEDVNDNPPEILQEY--VVICKPKMGYTDIL--AVDP 599

Qy    694 DQHLFRGPHFTFSLGSGS--LQNDWEVSKINGTHARLS-TRHTDFEERAYVVLIRINDGG 750
          |: :   | | ||| : |  :   | ::|:| | ||||  ::  |:|  | : | : |
Db    600 DEPVHGAP-FYFSLPNTSPEISRLWSLTKVNDTAARLSYQKNAGFQE--YTIPITVKD-- 654

Qy    751 RPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAV-GILLTTLLVIGIILAVVFIR 809
          |       | |  | |   :  |     ||:     |:  |||   |:  ::| :|
Db    655 RAGQAATKLLRVNLCECTHPTQCRATSRSTGVILGKWAILAILLGIALLFSVLLTLVCGV 714

Qy    810 IKKDKGK 816
            |||
Db    715 FGATKGK 721


RESULT 14
IJHUDB
desmocollin 3b precursor - human
N;Alternate names: desmosomal glycoprotein III
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 07-May-1999
C;Accession: A40390; S16464
R;Parker, A.E.; Wheeler, G.N.; Arnemann, J.; Pidsley, S.C.; Ataliotis, P.; Thomas, C.L.;
J. Biol. Chem. 266, 10438-10445, 1991
A;Title: Desmosomal glycoproteins II and III. Cadherin-like junctional molecules generat
A;Reference number: A40390; MUID:91244819; PMID:2037591
A;Accession: A40390
A;Molecule type: mRNA
A;Residues: 1-847 <PAR>
A;Cross-references: GB:X56807
A;Note: it is uncertain whether Met-1 is the initiator or whether translation is initiat
R;Buxton, R.S.; Cowin, P.; Franke, W.W.; Garrod, D.R.; Green, K.J.; King, I.A.; Koch, P.
J. Cell Biol. 121, 481-483, 1993
A;Title: Nomenclature of the desmosomal cadherins.
A;Reference number: A43032; MUID:93252984; PMID:8486729
A;Contents: annotation; nomenclature
C;Genetics:
A;Gene: GDB:DSC3; DSC2; DSC1; DS
A;Cross-references: GDB:126552; OMIM:600271
A;Map position: 18q12.1-18q12.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycoprot
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-135/Domain: propeptide #status predicted <PRO>
F;136-847/Product: desmocollin 3b #status predicted <MAT>
F;136-695/Domain: extracellular #status predicted <EXT>
F;138-243/Domain: cadherin repeat homology <CR1>
F;246-355/Domain: cadherin repeat homology <CR2>
F;358-471/Domain: cadherin repeat homology <CR3>
F;474-577/Domain: cadherin repeat homology <CR4>
F;578-680/Domain: cadherin repeat homology <CR5>
F;696-718/Domain: transmembrane #status predicted <TMM>
F;719-847/Domain: intracellular #status predicted <INT>
F;166,392,546,629/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             15.4%;  Score 667.5;  DB 1;  Length 847;
  Best Local Similarity   30.9%;  Pred. No. 6.2e-34;
  Matches  186;  Conservative  104;  Mismatches  253;  Indels   59;  Gaps   21;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQE- 289
          | || |  |:||| | |: : ||: :|    |::        ||:|  ||  | ::::
Db    137 W-APIPCSMLENSLGPFPLFLQQVQ-SDTAQNYTIYYSIRGPGVDQE--PRNLFYVERDT 192

Qy    290 GDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEV 349
          |::| |:|:|||: :::   | |    |     || : :|::| ||| |        | :
Db    193 GNLYCTRPVDREQYESFEIIAFATTPDGYTPELPLPLIIKIEDENDNYPIFTEETYTFTI 252

Qy    350 QENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQS 409
          || |:| ::| : | |:|| :| :: | | |: | |  |   || :    |::
Db    253 FENCRVGTTVGQVCATDKDEPDTMHTRLKYSIIGQVP--PSPTLFSMHPTTGVITTTSSQ 310

Qy    410 LKKQDTPQYNLTIEVSDKD-----FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTN 464
          | ::   :| | |:| | |      :|     ||: |:|| :| | :: |  :|  |:
Db    311 LDRELIDKYQLKIKVQDMDGQYFGLQTTSTCIINIDDVNDHLPTFTRTSY--VTSVEENT 368

Qy    465 IGSTILTIQATDADEPPTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFE-- 522
          :   || :   | |   | : :  | |:||:  |   : ||  || | : : |||::|
Db    369 VDVEILRVTVEDKDLVNTANWRANYTILKGNENGNFKIVTDAKTNEGVLCVVKPLNYEEK 428

Qy    523 ------TAAVSNIVFKAE-NPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKV 575
                |:   |  | :|         :| | |  |: | | :| |: :  :   ::
Db    429 QQMILQIGVVNEAPFSREASPR--------SAMSTATVTVNVEDQDEGPECNPPIQTVRM 480

Qy    576 SEDVAIGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA--- 629
          |:  :||      | |||      | |    |  ||: ||  || |     |||||
Db    481 KENAEVGTTSNGYKAYDPETRSSSGIRYKKLTDPTGWVTIDENTGSIKVFRSLDREAETI 540

Qy    630 -GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF 688
           | : |:|:: || : :      :|| ||||| | : |    :  | |  :   ::
Db    541 KNGIYNITVLASDQGGRTCTGT--LGIILQDVNDNSPFIPK--KTVIICKPTMSSAEIV- 595

Qy    689 EATDDDQHLFRGPHFTFSLGSGS--LQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRI 746
           | | |: :  || | ||| | :  :|  | :  || | |||| :: |    :||| | :
Db    596 -AVDPDEPI-HGPPFDFSLESSTSEVQRMWRLKAINDTAARLSYQN-DPPFGSYVVPITV 652

Qy    747 NDGGRPPLEGIVSLPVTFCSCV-EGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAV 805
           |  |  :  : || || | |: |  |      : |   | :    :|  || | :: :
Db    653 RD--RLGMSSVTSLDVTLCDCITENDCTHRVDPRIGGGGVQLGKWAILAILLGIALLFCI 710

Qy    806 VF 807
          :|
Db    711 LF 712


RESULT 15
IJHUDA
desmocollin 3a precursor - human
N;Alternate names: desmosomal glycoprotein II
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 07-May-1999
C;Accession: B40390; S16465
R;Parker, A.E.; Wheeler, G.N.; Arnemann, J.; Pidsley, S.C.; Ataliotis, P.; Thomas, C.
J. Biol. Chem. 266, 10438-10445, 1991
A;Title: Desmosomal glycoproteins II and III. Cadherin-like junctional molecules gene
A;Reference number: A40390; MUID:91244819; PMID:2037591
A;Accession: B40390
A;Molecule type: mRNA
A;Residues: 1-901 <PAR>
A;Cross-references: GB:X56807
A;Note: it is uncertain whether Met-1 is the initiator or whether translation is init
R;Buxton, R.S.; Cowin, P.; Franke, W.W.; Garrod, D.R.; Green, K.J.; King, I.A.; Koch,
J. Cell Biol. 121, 481-483, 1993
A;Title: Nomenclature of the desmosomal cadherins.
A;Reference number: A43032; MUID:93252984; PMID:8486729
A;Contents: annotation; nomenclature
C;Genetics:
A;Gene: GDB:DSC3; DSC2; DSC1; DS
A;Cross-references: GDB:126552; OMIM:600271
A;Map position: 18q12.1-18q12.1
C;Superfamily: cadherin; cadherin repeat homology
C;Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycop
F;1-28/Domain: signal sequence #status predicted <SIG>
F;29-135/Domain: propeptide #status predicted <PRO>
F;136-901/Product: desmocollin 3a #status predicted <MAT>
F;136-695/Domain: extracellular #status predicted <EXT>
F;138-243/Domain: cadherin repeat homology <CR1>
F;246-355/Domain: cadherin repeat homology <CR2>
F;358-471/Domain: cadherin repeat homology <CR3>
F;474-577/Domain: cadherin repeat homology <CR4>
F;578-680/Domain: cadherin repeat homology <CR5>
F;696-718/Domain: transmembrane #status predicted <TMM>
```

```
F;719-901/Domain: intracellular #status predicted <INT>
F;166,392,546,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;864/Binding site: phosphate (Ser) (covalent) #status predicted

  Query Match             15.4%;  Score 667.5;  DB 1;  Length 901;
  Best Local Similarity   30.9%;  Pred. No. 6.9e-34;
  Matches  186;  Conservative  104;  Mismatches  253;  Indels   59;  Gaps   21;

Qy     239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQE- 289
           | || |  |:||| | |: : ||: :|    |::        ||:|  ||  | ::::
Db     137 W-APIPCSMLENSLGPFPLFLQQVQ-SDTAQNYTIYYSIRGPGVDQE--PRNLFYVERDT 192

Qy     290 GDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEV 349
           |::| |:|:|||: :::   | |    |     || : :|::| ||| |        | :
Db     193 GNLYCTRPVDREQYESFEIIAFATTPDGYTPELPLPLIIKIEDENDNYPIFTEETYTFTI 252

Qy     350 QENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQS 409
            || |:| ::| : | |:|| :| :: | | |: | |  |   || :    |::
Db     253 FENCRVGTTVGQVCATDKDEPDTMHTRLKYSIIGQVP--PSPTLFSMHPTTGVITTTSSQ 310

Qy     410 LKKQDTPQYNLTIEVSDKD-----FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTN 464
           | ::   :| | |:| | |      :|     ||: |:|| :| | :: |  :|  |:
Db     311 LDRELIDKYQLKIKVQDMDGQYFGLQTTSTCIINIDDVNDHLPTFTRTSY--VTSVEENT 368

Qy     465 IGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFE-- 522
           :   || :   | |   | : :  | |:||:  |   : ||  || | : : |||::|
Db     369 VDVEILRVTVEDKDLVNTANWRANYTILKGNENGNFKIVTDAKTNEGVLCVVKPLNYEEK 428

Qy     523 ------TAAVSNIVFKAE-NPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKV 575
                 |:   |  | :|         :| | |  |: | | :| |: :  :   ::
Db     429 QQMILQIGVVNEAPFSREASPR--------SAMSTATVTVNVEDQDEGPECNPPIQTVRM 480

Qy     576 SEDVAIGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA--- 629
           |:  :||      | |||      | |    |  ||: ||  || |     |||||
Db     481 KENAEVGTTSNGYKAYDPETRSSSGIRYKKLTDPTGWVTIDENTGSIKVFRSLDREAETI 540

Qy     630 -GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF 688
           | : |:|:: || : :      :|| ||||| | : |    :  | |  :   ::
Db     541 KNGIYNITVLASDQGGRTCTGT--LGIILQDVNDNSPFIPK--KTVIICKPTMSSAEIV- 595

Qy     689 EATDDDQHLFRGPHFTFSLGSGS--LQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRI 746
           | | |: :  || | ||| | :  :|  | :  || | |||| :: |    :||| | :
Db     596 -AVDPDEPI-HGPPFDFSLESSTSEVQRMWRLKAINDTAARLSYQN-DPPFGSYVVPITV 652

Qy     747 NDGGRPPLEGIVSLPVTFCSCV-EGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAV 805
           |  |  :  : || || | |: |  |      : |   | :    :|  || | ::  :
Db     653 RD--RLGMSSVTSLDVTLCDCITENDCTHRVDPRIGGGGVQLGKWAILAILLGIALLFCI 710

Qy     806 VF 807
           :|
Db     711 LF 712


Search completed: January 15, 2003, 08:18:23
Job time : 66.7352 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:22 ; Search time 18.0457 Seconds
(without alignments)
1912.277 Million cell updates/sec

Title: US-09-079-819-178
Perfect score: 4321
Sequence: 1 MILQAHLHSLCLLMLYLATG.....DKGKDNVESAQASEVKPLRS 832

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4321	100.0	832	1	CADH_HUMAN	Q12864 homo sapien
2	3450	79.8	827	1	CADH_MOUSE	Q9r100 mus musculu
3	3414	79.0	827	1	CADH_RAT	P55281 rattus norv
4	954	22.1	829	1	CADG_RABIT	Q28634 oryctolagus
5	922.5	21.3	829	1	CADG_HUMAN	O75309 homo sapien
6	877.5	20.3	830	1	CADG_MOUSE	O88338 mus musculu
7	715	16.5	913	1	CAD4_CHICK	P24503 gallus gall
8	705	16.3	913	1	CAD4_MOUSE	P39038 mus musculu
9	699.5	16.2	906	1	CAD2_HUMAN	P19022 homo sapien
10	696	16.1	912	1	CAD2_CHICK	P10288 gallus gall
11	694.5	16.1	906	1	CAD2_MOUSE	P15116 mus musculu
12	688.5	15.9	877	1	CAD2_BOVIN	P19534 bos taurus
13	686.5	15.9	906	1	CAD2_RAT	Q9z1y3 rattus norv
14	684.5	15.8	916	1	CAD4_HUMAN	P55283 homo sapien
15	674.5	15.6	902	1	DSC2_MOUSE	P55292 mus musculu
16	671.5	15.5	896	1	DSC3_HUMAN	Q14574 homo sapien
17	667.5	15.4	901	1	DSC2_HUMAN	Q02487 homo sapien
18	663.5	15.4	906	1	CADO_XENLA	P33147 xenopus lae
19	661	15.3	896	1	DSC3_BOVIN	Q28060 bos taurus
20	657	15.2	713	1	CADD_HUMAN	P55290 homo sapien
21	653.5	15.1	863	1	DSC2_BOVIN	P33545 bos taurus
22	647.5	15.0	905	1	CADN_XENLA	P20310 xenopus lae
23	647	15.0	714	1	CADD_MOUSE	Q9wtr5 mus musculu
24	642	14.9	895	1	DSC3_MOUSE	P55850 mus musculu
25	634	14.7	887	1	CAD1_CHICK	P08641 gallus gall
26	630.5	14.6	712	1	CADD_CHICK	P33150 gallus gall
27	609	14.1	886	1	CAD1_RAT	Q9r0t4 rattus norv
28	608	14.1	790	1	CADI_HUMAN	Q13634 homo sapien
29	605	14.0	4590	1	FATH_HUMAN	Q14517 homo sapien
30	604	14.0	884	1	CAD1_MOUSE	P09803 mus musculu
31	601.5	13.9	3579	1	STAN_DROME	Q9v5n8 drosophila
32	597.5	13.8	732	1	CADL_CHICK	P33145 gallus gall
33	595.5	13.8	886	1	DSC1_MOUSE	P55849 mus musculu
34	591	13.7	882	1	CAD1_HUMAN	P12830 homo sapien
35	590	13.7	872	1	CAD1_XENLA	P30944 xenopus lae
36	587	13.6	784	1	CADF_MOUSE	P33146 mus musculu
37	585.5	13.6	893	1	DSC1_BOVIN	Q01107 bos taurus
38	580.5	13.4	880	1	CADF_XENLA	P33148 xenopus lae
39	576	13.3	3014	1	CLR1_HUMAN	Q9nyq6 homo sapien
40	574.5	13.3	796	1	CADB_MOUSE	P55288 mus musculu
41	574.5	13.3	894	1	DSC1_HUMAN	Q08554 homo sapien
42	571.5	13.2	789	1	CADA_CHICK	P79995 gallus gall
43	571	13.2	3034	1	CLR1_MOUSE	O35161 mus musculu
44	567.5	13.1	788	1	CADA_HUMAN	Q9y6n8 homo sapien
45	565	13.1	814	1	CADF_HUMAN	P55291 homo sapien

ALIGNMENTS

RESULT 1
CADH_HUMAN
ID CADH_HUMAN STANDARD; PRT; 832 AA.
AC Q12864; Q15336;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-17 precursor (Liver-intestine-cadherin) (LI-cadherin)
DE (Intestinal peptide-associated transporter HPT-1).
GN CDH17.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC TISSUE=Colon adenocarcinoma;
RX MEDLINE=94204643; PubMed=8153632;
RA Dantzig A.H., Hoskins J., Tabas L.B., Bright S., Shepard R.L.,
RA Jenkins I.L., Duckworth D.C., Sportsman J.R., MacKensen D.,
RA Rosteck P.R. Jr., Skatrud P.L.;
RT "Association of intestinal peptide transport with a protein related to
RT the cadherin superfamily.";
RL Science 264:430-433(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Boettinger A., Kreft B., Fieger C., Dlouhy B., Berndorff D.,
RA Goessner R., Tauber R.;
RT "Molecular cloning of human LI-cadherin: evidence for a novel type
RT of cadherin within the cadherin superfamily.";
RL Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. LI-CADHERIN MAY HAVE A ROLE
CC IN THE MORPHOLOGICAL ORGANIZATION OF LIVER AND INTESTINE.
CC INVOLVED IN INTESTINAL PEPTIDE TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE GASTROINTESTINAL TRACT AND
CC PANCREATIC DUCT. NOT DETECTED IN KIDNEY, LUNG, LIVER, BRAIN,
CC ADRENAL GLAND AND SKIN.
CC -!- SIMILARITY: CONTAINS 7 CADHERIN DOMAINS.
CC ---

CC ---
DR EMBL; U07969; AAA19021.1; -.
DR EMBL; X83228; CAA58231.1; -.
DR HSSP; P15116; 1NCJ.
DR Genew; HGNC:1756; CDH17.
DR MIM; 603017; -.

```
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 7.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 6.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal; Transport.
FT   SIGNAL        1     22       POTENTIAL.
FT   CHAIN        23    832       CADHERIN-17.
FT   DOMAIN       23    787       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    788    808       POTENTIAL.
FT   DOMAIN      809    832       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       30    128       CADHERIN 1.
FT   DOMAIN      129    244       CADHERIN 2.
FT   DOMAIN      245    340       CADHERIN 3.
FT   DOMAIN      341    449       CADHERIN 4.
FT   DOMAIN      450    566       CADHERIN 5.
FT   DOMAIN      567    667       CADHERIN 6.
FT   DOMAIN      668    777       CADHERIN 7.
FT   CARBOHYD    149    149       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    250    250       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    546    546       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    587    587       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    722    722       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    115    115       E -> K (IN REF. 2).
FT   CONFLICT    446    446       I -> T (IN REF. 2).
FT   CONFLICT    734    734       D -> E (IN REF. 2).
FT   CONFLICT    739    739       A -> E (IN REF. 2).
SQ   SEQUENCE   832 AA;  92147 MW;  C6DB6DE60E8A1277 CRC64;

  Query Match             100.0%;  Score 4321;  DB 1;  Length 832;
  Best Local Similarity   100.0%;  Pred. No. 7.1e-256;
  Matches  832;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MILQAHLHSLCLLMLYLATGYGQQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MILQAHLHSLCLLMLYLATGYGQQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTF 60

Qy      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ELTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIN 120

Qy     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 DNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYF 180

Qy     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 QINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWK 240

Qy     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 APKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDR 300

Qy     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 EEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIG 360

Qy     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNL 420

Qy     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TIEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEP 480

Qy     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 FTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLV 540

Qy     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 FGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISY 600

Qy     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 SLRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDV 660

Qy     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 NDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSK 720

Qy     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 INGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQT 780

Qy     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832


RESULT 2
CADH_MOUSE
ID   CADH_MOUSE     STANDARD;      PRT;   827 AA.
AC   Q9R100;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cadherin-17 precursor (Liver-intestine-cadherin) (LI-cadherin) (BILL-
DE   cadherin) (P130).
GN   CDH17.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss Webster; TISSUE=Intestine;
RA   Angres B., Kim L., Tauber R.;
RT   "LI-cadherin gene expression during intestinal development.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., SEQUENCE OF 26-33; 52-58; 74-81; 117-123 AND
RP   490-509, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC   STRAIN=BALB/c; TISSUE=Fetal liver;
RX   MEDLINE=20469471; PubMed=10906147;
RA   Ohnishi K., Shimizu T., Karasuyama H., Melchers F.;
RT   "The identification of a nonclassical cadherin expressed during B cell
RT   development and its interaction with surrogate light chain.";
RL   J. Biol. Chem. 275:31134-31144(2000).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. LI-CADHERIN MAY HAVE A ROLE
CC       IN THE MORPHOLOGICAL ORGANIZATION OF LIVER AND INTESTINE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IS FOUND IN INTESTINE WITH
CC       LOWER EXPRESSION IN SPLEEN, BONE MARROW, LUNG AND TESTIS. NO
CC       EXPRESSION DETECTED IN LIVER, KIDNEY, HEART, BRAIN OR SKELETAL
CC       MUSCLE. EXPRESSED IN PRECURSOR B-CELLS AND MYELOID CELLS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION INCREASES IN PRO- AND PRE-B-I
CC       CELLS, DECREASES IN LARGE AND SMALL PRE-B-II CELLS, AND INCREASES
CC       AGAIN IN IMMATURE AND MATURE B-CELLS.
CC   -!- SIMILARITY: CONTAINS 7 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AF177669; AAD51125.1; -.
DR   EMBL; D87912; BAB03264.1; -.
DR   HSSP; P15116; 1NCI.
DR   MGD; MGI:1095414; Cdh17.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 7.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal; Transport.
FT   SIGNAL        1     25
FT   CHAIN        26    827       CADHERIN-17.
FT   DOMAIN       26    786       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    787    807       POTENTIAL.
FT   DOMAIN       808    827       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        29    127       CADHERIN 1.
FT   DOMAIN       128    243       CADHERIN 2.
FT   DOMAIN       244    339       CADHERIN 3.
FT   DOMAIN       340    448       CADHERIN 4.
FT   DOMAIN       449    565       CADHERIN 5.
FT   DOMAIN       566    666       CADHERIN 6.
FT   DOMAIN       667    776       CADHERIN 7.
FT   CARBOHYD     148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     249    249       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     586    586       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   827 AA;  91645 MW;  CDCECEA5A76E2B58 CRC64;

  Query Match             79.8%;  Score 3450;  DB 1;  Length 827;
  Best Local Similarity   79.2%;  Pred. No. 8.5e-203;
  Matches  658;  Conservative   75;  Mismatches   94;  Indels    4;  Gaps    1;

Qy       2 ILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFE 61
           :: | || |||| |||  |||:||||||||||||||:|||||||:||||| ||||||||
Db       1 MVSAQLHFLCLLTLYLTCGYGEEGKFSGPLKPMTFSIFEGQEPSQVIFQFKTNPPAVTFE 60

Qy      62 LTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIIEVKDIND 121
           |||||| || ||::|||: |||||||: |:||:||||::| ||:|||||||||||||||
Db      61 LTGETDGIFKIEKDGLLYHTRALDRETRAVHHLQLAALDSHGAIVDGPVPITIIEVKDIND 120

Qy     122 NRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQ 181
           ||||||||||||||||||||||:||||||||||||||||:|||||||| ||:|||||
Db     121 NRPTFLQSKYEGSVRQNSRPGKPFMYVNATDLDDPATPNGQLFYQIVIQLPQINDVMYFQ 180

Qy     182 INNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKA 241
           |::|||||||| ||||||:| ||||||||:|||||||||||||||||| ||| : ||||||
Db     181 IDSKTGAISLTPEGSQELDPVKNPSYNLVVSVKDMGGQSENSFSDTTYVDISIRENIWKA 240

Qy     242 PKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDRE 301
           |:|||: |||||||||||||:||||||||||:|||| |||||||||:||||| ||||
Db     241 PEPVEIRENSTDPHPIKITQVQWNDPGAQYSLVNKEKLSPFPFSIDQEGNIYVTQALDRE 300

Qy     302 EKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGT 361
           ||:::||:| |||| |||||:|||||:||| |||||||| |||||||||| ||||||
Db     301 EKNSHVFFATAKDENGKPLAYPLEIYVKVIDINDNPPTCLSPVTVFEVQENEPLGNSIGI 360

Qy     362 LTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLT 421
             ||| || |  || | |::|:||||:| ||||||  | | :||:||||||||:||||||:
Db     361 FEAHDMDEANNINSILKYKLVDQTPKVPSDGLFLIGEYEGKVQLSKQSLKKQDSPQYNLS 420

Qy     422 IEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPF 481
           ||||| ||||||::|:||||||||||||| |:||: ||:||| |||||| ||||||||||
Db     421 IEVSDVDFKTLCYIQVNVIDINDQIPIFETSNYGSKTLSEDTAIGSTILIIQATDADEPF 480

Qy     482 TGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVF 541
           |||||||| |::||:|||| | ||| || ||| ||||||||| ||:|||:|||||||||
Db     481 TGSSKILYKIVQGDTEGRLEVVTDPTTNAGYVKIKKPLDFETQPVSSIVFQAENPEPLVK 540

Qy     542 GVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYS 601
           |::|||||||| | |||||||| | | | :||| |||| |:|::|||||||:|||| :|||
Db     541 GIEYNASSFASFELIVTDVNEVPVFPQRIFQANVSEDAAVGSRVGNVTARDPEGLTVSYS 600

Qy     602 LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVN 661
           |:|: ||||||| ||||||| ||||||  | ||||||||||||||||| ::|||:| |||
Db     601 LKGNMRGWLKIDSVTGEIFSAAPLDRETESVYRVQVVATEVGGSSLSSTADFHLVLTDVN 660

Qy     662 DNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKI 721
           |||||||||||||||||||||||||||| |||||   | | |||:||   ||:|||||||
Db     661 DNPPRLAKDYTGLFFCHPLSAPGSLIFEVTDDDQQSLRRPKFTFALGREGLQSDWEVSKI 720

Qy     722 NGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTG 781
           |||||||||||| ||| | : ||||||||:||:|| | |||||| |||||||||||| | |
Db     721 NGTHARLSTRHTRFEEQVYNIPIRINDGGQPPMEGTVFLPVTFCQCVEGSCFRPAGRQDG 780

Qy     782 IPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           |||||||||||||| ||||||||||||::|||    ||: |: | |||||
Db     781 IPTVGMAVGILLTTFLVIGIILAVVFIRMRKDK----VENPQSPENKPLRS 827


RESULT 3
CADH_RAT
ID   CADH_RAT       STANDARD;      PRT;   827 AA.
AC   P55281;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cadherin-17 precursor (Liver-intestine-cadherin) (LI-cadherin).
GN   CDH17.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Liver;
RX   MEDLINE=94266966; PubMed=8207063;
RA   Berndorff D., Gessner R., Kreft B., Schnoy N., Lajous-Petter A.-M.,
RA   Loch N., Reutter W., Hortsch M., Tauber R.;
RT   "Liver-intestine cadherin: molecular cloning and characterization of
RT   a novel Ca(2+)-dependent cell adhesion molecule expressed in liver
RT   and intestine.";
RL   J. Cell Biol. 125:1353-1369(1994).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. LI-CADHERIN MAY HAVE A ROLE
CC       IN THE MORPHOLOGICAL ORGANIZATION OF LIVER AND INTESTINE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: LIVER AND INTESTINE.
CC   -!- SIMILARITY: CONTAINS 7 CADHERIN DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X78997; CAA55631.1; -.
DR   HSSP; P15116; 1NCI.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 7.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 6.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 6.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     21       POTENTIAL.
```

```
FT   CHAIN        22    827       CADHERIN-17.
FT   DOMAIN       22    786       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    787    807       POTENTIAL.
FT   DOMAIN       808    827       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        29    127       CADHERIN 1.
FT   DOMAIN       128    243       CADHERIN 2.
FT   DOMAIN       244    339       CADHERIN 3.
FT   DOMAIN       340    448       CADHERIN 4.
FT   DOMAIN       449    565       CADHERIN 5.
FT   DOMAIN       566    666       CADHERIN 6.
FT   DOMAIN       667    776       CADHERIN 7.
FT   CARBOHYD     148    148       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     183    183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     249    249       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     418    418       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     545    545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     573    573       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   827 AA;  91862 MW;  E1A0B03646562C01 CRC64;

  Query Match           79.0%;  Score 3414;  DB 1;  Length 827;
  Best Local Similarity   78.8%;  Pred. No. 1.3e-200;
  Matches  655;  Conservative   70;  Mismatches  102;  Indels    4;  Gaps    1;

Qy       2 ILQAHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQFKANPPAVTFE 61
           :: | || |||| |||   ||||||||||||||||||:|||||||||||||||||||||
Db       1 MVSAQLHFLCLLTLYLTGAYGQEGKFSGPLKPMTFSIFEGQEPSQIIFQFKANPPAVTFE 60

Qy      62 LTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIND 121
           |||||| || ||::|||| | ||||||: |:||:|||| | ||:|||| ||||||||
Db      61 LTGETDGIFKIEKDGLLYHTRVLDRETRAVHHLQLAALDSQGAIVDGPVPIIIEVKDIND 120

Qy     122 NRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQ 181
           |||||||:|||||||||||||||:|||||||||||||||:|||||||| ||||||||
Db     121 NRPTFLQTKYEGSVRQNSRPGKPFMYVNATDLDDPATPNGQLFYQIVIQLPKINNVMYFQ 180

Qy     182 INNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKA 241
           |:|||||||||| |||| |:| ||| ||||:||||||||:||||||||||| | ||||||
Db     181 IDNKTGAISLTPEGSQVLDPIKNPYYNLVVSVKDMGGQNENSFSDTTSVDITVRENIWKA 240

Qy     242 PKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDRE 301
           |:|||: || |||||||||||:|||||| |||::|||||:|||||||||:|||||||||
Db     241 PEPVEIRENLTDPHPIKITQVQWNDPGAHYSLINKEKLPQFPFSIDQEGNIYVTQPLDRE 300

Qy     302 EKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGT 361
           |||::||:| |||| ||||:|||| ||| |||||||| | |||||||||| ||:|||
Db     301 EKDSHVFFATAKDENGKPLAYPLEIRVKVIDINDNPPTCLSQVTVFEVQENEVLGSSIGI 360

Qy     362 LTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLT 421
             ||| || |  || | ||:|:||||:| | ||||  | | :|| |:||||||:||||||
Db     361 FAAHDMDEANNINSILKYRLVDQTPKVPSDELFLIDEYGGKVQLGKRSLKKQDSPQYNLT 420

Qy     422 IEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPF 481
           :|||| |||||| :|:|||||||||||||:||||: ||:||| |||||| |||||  ||||
Db     421 VEVSDIDFKTLCSLQVNVIDINDQIPIFERSDYGSKTLSEDTAIGSTILIIQATDDDEPF 480

Qy     482 TGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVF 541
           |||||||| |::||:|||| | ||| |||||| |:|||||  |::||||||||||||
Db     481 TGSSKILYKIVQGDTEGRLEVVTDPMTNTGYVKIRKPLDFETEPVTSIVFKAENPEPLVN 540

Qy     542 GVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYS 601
           |::||||||| | | |||||| | | | :||| |||| ||||||| |||:|||||  :|||
Db     541 GIEYNASSFASFELTVTDVNEVPVFPQQIFQANVSEDTAIGTKVGTVTARDPEGLTVSYS 600

Qy     602 LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVN 661
           |: : ||||||| |||:||| ||||||  | ||||||||||||||||| : |||:||||||
Db     601 LKDNKRGWLKIDSVTGDIFSTAPLDRETESVYRVQVVATEVGGSSLSSTAYFHLVLMDVN 660

Qy     662 DNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKI 721
           |||||||||||||||||||||||||||||||||  | | |||:||  |||:||||||||
Db     661 DNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQQSVRRPKFTFALGRESLQSDWEVSKI 720

Qy     722 NGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTG 781
           |||||||||:|| |||: | : | |||||:||:|||||| |||| ||:||||||||:| |
Db     721 NGTHARLSTKHTRFEEQVYDIPILINDGGQPPMEGIVSLSVTFCQCVDGSCFRPAGNQVG 780

Qy     782 IPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPLRS 832
           |||||||||||||| ||||||||||||::|||    ||  |: | |||||
Db     781 IPTVGMAVGILLTTFLVIGIILAVVFIRMRKDK----VEDPQSPENKPLRS 827


RESULT 4
CADG_RABIT
ID   CADG_RABIT     STANDARD;      PRT;   829 AA.
AC   Q28634;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cadherin-16 precursor (Kidney-specific cadherin) (Ksp-cadherin).
GN   CDH16.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND OF SEQUENCE 570-586.
RC   STRAIN=New Zealand white;
RX   MEDLINE=95340560; PubMed=7615566;
RA   Thomson R.B., Igarashi P., Biemesderfer D., Kim R., Abu-Alfa A.,
RA   Soleimani M., Aronson P.S.;
RT   "Isolation and cDNA cloning of Ksp-cadherin, a novel kidney-specific
RT   member of the cadherin multigene family.";
RL   J. Biol. Chem. 270:17594-17601(1995).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC. LIMITED TO THE BASOLATERAL
CC       MEMBRANES OF RENAL TUBULAR EPITHELIAL CELLS.
CC   -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U28945; AAC48472.1; -.
DR   HSSP; P09803; 1SUH.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 6.
DR   SMART; SM00112; CA; 6.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 6.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    829       CADHERIN-16.
FT   DOMAIN       19    786       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    787    807       POTENTIAL.
FT   DOMAIN       808    829       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        25    126       CADHERIN 1.
FT   DOMAIN       131    235       CADHERIN 2.
FT   DOMAIN       242    336       CADHERIN 3.
FT   DOMAIN       341    449       CADHERIN 4.
FT   DOMAIN       455    564       CADHERIN 5.
FT   DOMAIN       569    665       CADHERIN 6.
FT   DOMAIN       666    786       ECTODOMAIN G.
FT   CARBOHYD     517    517       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     602    602       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     709    709       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     722    722       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   829 AA;  88827 MW;  D2DF10E6C47A43B9 CRC64;
```

```
  Query Match             22.1%;  Score 954;  DB 1;  Length 829;
  Best Local Similarity   31.6%;  Pred. No. 1.1e-50;
  Matches  251;  Conservative  142;  Mismatches  354;  Indels   48;  Gaps   22;

Qy      62 LTGET----DNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|::    :  | :| | | |   |||||| :: : :||     :| :: ||  :|: |
Db      57 LSGDSGVAAEGPFSVEAESGFLLVTRALDREEQAEYQIQVTLEAEDGHVLWGPQSVTVHV 116

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : | :||| |||:: |:| |:| | |  | : |: | |   :
Db     117 KDENDQVPQFSQALYSARLSQGTRPGVPFLFLEASDGDEPGTANSDLRFHILSQTPAQPS 176

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
           |::  : ||::|: |||   : |    | |:: ||||| |: :    | :|:| : |
Db     177 PDVFRLEPRLGALALSPEGSAGFDHALEGPYQLLVQVKDMGDQA-SGHQATATVEISIVE 235

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           : |   :|| : ||   |:|  : || |:     | |   |  | || :| || :|||
Db     236 STWTPLEPVHLAENLKVPYPHHLAQVHWSGGDVHYRL---ESQPPGPFDVDTEGKLYVTG 292

Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCP---SPVTVFEVQENE 353
           |||| :: ||    |::  |:  : |||:|| | | ||: | ||    ||:| |:
Db     293 ELDREAQEQYVLQVQAQNSRGEDYAEPLELHVVVTDENDHAPVCPPRGPPVSVPELSPP- 351

Qy     354 RLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKL-PMDGLFLIQTYAGMLQLAKQSLKK 412
             | :: ||:| | |   : || : ||::   |:  |  | | :   :| : |    |:
Db     352 --GTAVTTLSAEDADAPGSPNSHVVYRLLSPEPQEGPEGGAFQLDPTSGSVSLGAAPLEA 409

Qy     413 QDTPQYNLTIEV-------SDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNI 465
                |: ::|       ::    : | | : | |:||  | |  |  | ::| |||
Db     410 GQ----NMLLQVLAVDLAGAEAGLSSTCEVAVTVTDVNDHAPEFTSSQVGPVSLPEDTEP 465

Qy     466 GSTILTIQATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFET 523
           |: : |: ||||| ||   : ::: : |  || ||  |:| :| :    : : | | :|
Db     466 GTLVATLTATDADLEP---AFRLMDFTIEAGDGEGTFGLDWEPDSGHVQLYLLKNLSYEA 522

Qy     524 AAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGT 583
           |    :|    |   | |     : |  |::|  |   |:  |  ::| :  :   |:
Db     523 APSHTVVVVVRNVVETV-GPGPGPGATATVTVLVEKVMPPPRLEQKSYEADIPVNAPAGS 581

Qy     584 KVGNVTAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEV 642
            :  :   :|    : :||  |: ||  |  |:||: :  ||      |  | | | | :
Db     582 FLLTIQPAEPWNGALRFSLVNDSEGWFCIQKVSGEVHTARPLQGARPGDSYTVLVEAQDA 641

Qy     643 GGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDD-DQHLFRGP 701
               ||: :   :  :       |           | |    | ||   ::| |     ||
Db     642 DAPRLSTSAALVIHFLRAPPAPALPLAPMPSRHLCTPRQDHGVLIPAPSEDPDMATGHGP 701

Qy     702 HFTFSLG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSL 760
           ::|:|| : ::| || :  :| :|| |:      | | ::| : ::   |     :  |
Db     702 -YSFALGPNPTVQRDWRLQPLNDSHAFLTLALHWVEPREHIVPVVVSQDAR-----VWQL 755

Qy     761 P--VTFCSC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKD 817
           |  |  | |  || | |  |   |:||   ||||: ||  ||  | ::|  :   : ||
Db     756 PVRVVVCRCNTEGECMRKVGRMKGMPTKLSAVGILVGTLAAIGFFLILIFTHLALARKKD 815

Qy     818 NVESAQASEVKPLRS 832
             | |  | ||::
Db     816 --LDAPADNV-PLKA 827


RESULT 5
CADG_HUMAN
ID   CADG_HUMAN     STANDARD;      PRT;   829 AA.
AC   O75309;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cadherin-16 precursor (Kidney-specific cadherin) (Ksp-cadherin).
GN   CDH16.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98389630; PubMed=9721215;
RA   Thomson R.B., Ward D.C., Quaggin S.E., Igarashi P., Muckler Z.E.,
RA   Aronson P.S.;
RT   "cDNA cloning and chromosomal localization of the human and mouse
RT   isoforms of Ksp-cadherin.";
RL   Genomics 51:445-451(1998).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
CC   -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF016272; AAC34255.1; -.
DR   HSSP; P09803; 1SUH.
DR   Genew; HGNC:1755; CDH16.
DR   MIM; 603118; -.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 6.
DR   SMART; SM00112; CA; 6.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    829       CADHERIN-16.
FT   DOMAIN       19    786       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    787    807       POTENTIAL.
FT   DOMAIN      808    829       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       25    126       CADHERIN 1.
FT   DOMAIN      131    235       CADHERIN 2.
FT   DOMAIN      242    336       CADHERIN 3.
FT   DOMAIN      341    449       CADHERIN 4.
FT   DOMAIN      455    564       CADHERIN 5.
FT   DOMAIN      569    665       CADHERIN 6.
FT   DOMAIN      666    786       ECTODOMAIN G.
FT   CARBOHYD    517    517       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    602    602       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    722    722       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   829 AA;  89922 MW;  41966915E4BFC891 CRC64;

  Query Match             21.3%;  Score 922.5;  DB 1;  Length 829;
  Best Local Similarity   30.2%;  Pred. No. 9.2e-49;
  Matches  237;  Conservative  163;  Mismatches  353;  Indels   31;  Gaps   19;

Qy      62 LTGE----TDNIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:    |:  | :: : | |   ||||||:: : |||     :| :: || |: : |
Db      57 LSGDSGKATEGPFAMDPDSGFLLVTRALDREEQAEYQLQVTLEMQDGHVLWGPQPVLVHV 116

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : : :||| |||:: |:| |:| | |  | : |: | |   :
Db     117 KDENDQVPHFSQAIYRARLSRGTRPGIPFLFLEASDRDEPGTANSDLRFHILSQAPAQPS 176

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
             ||:  : ||::|: :||  |: |   :| |:: ||||| |: :    | :|:: : |
Db     177 PDMFQLEPRLGALALSPKGSTSLDHALERTYQLLVQVKDMGDQA-SGHQATATVEVSIIE 235

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           : | : :|: : ||    :|  : || |:     | |   |  | || :: ||::|||:
Db     236 STWVSLEPIHLAENLKVLYPHHMAQVHWSGGDVHYHL---ESHPPGPFEVNAEGNLYVTR 292
```

```
Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
           ||||  :  |:    |:: :|:  : |||:|| | | ||| | ||       : |    |
Db     293 ELDREAQAEYLLQVRAQNSHGEDYAAPLELHVLVMDENDNVPICPPRDPTVSIPELSPPG 352

Qy     357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQLAKQSLKK-QD 414
             :  |:| | |   : || : |:::   |:  ::|  | :  :| : |    |:  |:
Db     353 TEVTRLSAEDADAPGSPNSHVVYQLLSPEPEDGVEGRAFQVDPTSGSVTLGVLPLRAGQN 412

Qy     415 TPQYNLTIEVSDKD--FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTI 472
               | ::::  :  | : | |:: | ||||  | |  |  | ::| ||   |: :  :
Db     413 ILLLVLAMDLAGAEGGFSSTCEVEVAVTDINDHAPEFITSQIGPISLPEDVEPGTLVAML 472

Qy     473 QATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIK--KPLDFETAAVSN 528
           | ||| ||   : ::: : | :||:||  |:| :|  ::|:| ::  | | :| |
Db     473 TAIDADLEP---AFRLMDFAIERGDTEGTFGLDWEP--DSGHVRLRLCKNLSYEAAPSHE 527

Qy     529 IVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNV 588
           :|   ::   || |      : | |:::|  |  |: |  ::| | |      |: :  :
Db     528 VVVVVQSVAKLV-GPGPGPGATATVTVLVERVMPPPKLDQESYEASVPISAPAGSFLLTI 586

Qy     589 TAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVATEVGGSSL 647
              ||    : :||  |: ||| |:  :||: :   |   : |  | | | | :     |
Db     587 QPSDPISRTLRFSLVNDSEGWLCIEKFSGEVHTAQSLQGAQPGDTYTVLVEAQDTDEPRL 646

Qy     648 SSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIF-EATDDDQHLFRGPHFTFS 706
           |: :   :  :     |          : | |    | ::   : | |     || ::|:
Db     647 SASAPLVIHFLKAPPAPALTLAPVPSQYLCTPRQDHGLIVSGPSKDPDLASGHGP-YSFT 705

Qy     707 LG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFC 765
           || : ::| || :  :||:|| |:      | | ::: : ::   :      : : | |
Db     706 LGPNPTVQRDWRLQTLNGSHAYLTLALHWVEPREHIIPVVVSHNAQ---MWQLLVRVIVC 762

Qy     766 SC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQA 824
           | ||| | |  |   |:||  |||||: ||: ||| | ::|     : ||  : | :
Db     763 RCNVEGQCMRKVGRMKGMPTKLSAVGILVGTLVAIGIFLILIFTHWTMSRKKDPDQPADS 822

Qy     825 SEVK 828
             :|
Db     823 VPLK 826
```

```
RESULT 6
CADG_MOUSE
ID   CADG_MOUSE     STANDARD;      PRT;   830 AA.
AC   O88338; Q9JLZ5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cadherin-16 precursor (Kidney-specific cadherin) (Ksp-cadherin).
GN   CDH16.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RX   MEDLINE=98389630; PubMed=9721215;
RA   Thomson R.B., Ward D.C., Quaggin S.E., Igarashi P., Muckler Z.E.,
RA   Aronson P.S.;
RT   "cDNA cloning and chromosomal localization of the human and mouse
RT   isoforms of Ksp-cadherin.";
RL   Genomics 51:445-451(1998).
RN   [2]
RP   SEQUENCE OF 1-25 FROM N.A.
RC   STRAIN=129/Sv;
RA   Whyte D.A., Thomson R.B., Nix S.L., Zanjani R., Li C., Karp S.L.,
RA   Aronson P.S., Igarashi P.;
RT   "Ksp-cadherin gene promoter. I. Sequence analysis and renal
RT   epithelial-cell-specific activity.";
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC   -!- TISSUE SPECIFICITY: KIDNEY-SPECIFIC.
CC   -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF016271; AAC34254.1; -.
DR   EMBL; AF118228; AAF28836.1; -.
DR   HSSP; P09803; 1SUH.
DR   MGD; MGI:106671; Cdh16.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 6.
DR   SMART; SM00112; CA; 6.
DR   PROSITE; PS00232; CADHERIN_1; 1.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22    830       CADHERIN-16.
FT   DOMAIN       22    788       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    789    809       POTENTIAL.
FT   DOMAIN      810    830       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       27    128       CADHERIN 1.
FT   DOMAIN      133    237       CADHERIN 2.
FT   DOMAIN      244    338       CADHERIN 3.
FT   DOMAIN      343    451       CADHERIN 4.
FT   DOMAIN      457    566       CADHERIN 5.
FT   DOMAIN      571    667       CADHERIN 6.
FT   DOMAIN      668    788       ECTODOMAIN G.
FT   CARBOHYD    519    519       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    604    604       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    724    724       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   830 AA;  89859 MW;  5EB1F06B5B7C2F25 CRC64;

  Query Match             20.3%;  Score 877.5;  DB 1;  Length 830;
  Best Local Similarity   30.5%;  Pred. No. 5.1e-46;
  Matches  245;  Conservative  141;  Mismatches  353;  Indels   63;  Gaps   22;

Qy      62 LTGETD----NIFVIERE-GLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEV 116
           |:|:::    | | :: : | |   | |||| :: : |||     :| |: ||  :|: |
Db      59 LSGDSNTADQNTFAVDTDSGFLVATRTLDREEKAEYQLQVTLESEDGRILWGPQLVTVHV 118

Qy     117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
           || ||  | | |: |   : | :||| |||:: |:| | | | |  | : |: | |
Db     119 KDENDQVPQFSQAIYRAQLSQGTRPGVPFLFLEASDGDAPGTANSDLRFHILSQSPPQPL 178

Qy     177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE 236
             ||::   ||::|:  ||  |: |   :| |:: ||||| |     :  | |:| : |
Db     179 PDMFQLDPHLGALALSPSGSTSLDHALEETYQLLVQVKDMGDQPSGHQAIAT-VEISIVE 237

Qy     237 NIWKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQ 296
           | |   :|| : ||    :|  | || |:     | |   |  | | || :| || ::||
Db     238 NSWAPLEPVHLAENLKVVYPHSIAQVHWSGGDVHYQL---ESQPPGPFDVDTEGMLHVTM 294

Qy     297 PLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLG 356
           ||||  :  |     |:: :|:  : |||: | | | ||| | |        : |    |
Db     295 ELDREAQAEYQLQVRAQNSHGEDYAEPLELQVVVMDENDNAPVCSPHDPTVNIPELSPPG 354

Qy     357 NSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDG-LFLIQTYAGMLQL------AKQS 409
             |  |:| | |   : || : |:::   |:   :   | :   :| : |      | ||
Db     355 TEIARLSAEDLDAPGSPNSHIVYQLLSPEPEEGAENKAFELDPTSGSVTLGTAPLHAGQS 414

Qy     410 LKKQDTPQYNLTIEV--SDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGS 467
           :  |      | :::  |:    : | | : | |:|:  | |  |  | :|| ||   |:
```

```
Db    415 ILLQ-----VLAVDLAGSESGLSSTCEVTVMVTDVNNHAPEFINSQIGPVTLPEDVKPGA 469

Qy    468 TILTIQATDAD-EPFTGSSKIL-YHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAA 525
          : |: ||||| ||   : ::: : | :|| ||   :  :| ::   : ::| | :| |
Db    470 LVATLMATDADLEP---AFRLMDFAIEEGDPEGIFDLSWEPDSDHVQLRLRKNLSYEAAP 526

Qy    526 VSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKV 585
            :|    | | || |     :: |  |::|  |    :  |  ::  :      |: :
Db    527 DHKVVVVVSNIEELV-GPGPGPAATATVTILVERVVAPLKLDQESYETSIPVSTPAGSLL 585

Qy    586 GNVTAKDPEGLDISYSLRGDTRGWLKIDHVTGEIFSVAPLD-REAGSPYRVQVVA--TEV 642
            :   ||    : :||  |: ||| |  |:||: :   |   : |  | | | |  |:
Db    586 LTIQPSDPMSRTLRFSLVNDSEGWLCIKEVSGEVHTAQSLQGAQPGDTYTVLVEAQDTDK 645

Qy    643 GGSSLSSVSEFHLI------LMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQH 696
           | | |:    | :       : ::  | |          | |    | ::   ::|
Db    646 PGLSTSATVVIHFLKASPVPALTLSAGPSR--------HLCTPRQDYGVVVSGVSEDPDL 697

Qy    697 LFRGPHFTFSLG-SGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRI-NDGGRPPL 754
            |   ::|:|| : ::| || :  :| :|| |:      |   |:| : : :|     |
Db    698 ANRNGPYSFALGPNPTVQRDWRLQPLNDSHAYLTLALHWVEPGEYMVPVVVHHDTHMWQL 757

Qy    755 EGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFIRI--- 810
          :    : |  | | ||| | |  |   |:||   |||:|| ||  || || :|| :
Db    758 Q----VKVIVCRCNVEGQCMRKVGRMKGMPTKLSAVGVLLGTLAAIGFILILVFTHLALA 813

Qy    811 KKDKGKDNVESAQASEVKPLRS 832
          :||         | ::  ||::
Db    814 RKDLD-------QPADSVPLKA 828


RESULT 7
CAD4_CHICK
ID   CAD4_CHICK     STANDARD;      PRT;   913 AA.
AC   P24503;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
GN   CDH4.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Retina;
RX   MEDLINE=91299341; PubMed=1712604;
RA   Inuzuka H., Miyatani S., Takeichi M.;
RT   "R-cadherin: a novel Ca(2+)-dependent cell-cell adhesion molecule
RT   expressed in the retina.";
RL   Neuron 7:69-79(1991).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. MAY PLAY AN IMPORTANT ROLE
CC       IN RETINAL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EMBRYONIC BRAIN AND NEURONAL RETINA.
CC   -!- DEVELOPMENTAL STAGE: DETECTED ONLY AFTER SOME DEGREE OF NEURONAL
CC       DIFFERENTIATION HAS TAKEN PLACE AND PERSISTS AT LEAST UP TO THE
CC       NEWLY HATCHED STAGE.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D14459; BAA03356.1; -.
DR   PIR; JH0424; IJCHCR.
DR   HSSP; P15116; 1NCI.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     19       POTENTIAL.
FT   PROPEP       20    166       POTENTIAL.
FT   CHAIN       167    913       CADHERIN-4.
FT   DOMAIN      167    731       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    732    753       POTENTIAL.
FT   DOMAIN      754    913       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      167    274       CADHERIN 1.
FT   DOMAIN      275    389       CADHERIN 2.
FT   DOMAIN      390    504       CADHERIN 3.
FT   DOMAIN      505    610       CADHERIN 4.
FT   DOMAIN      611    721       CADHERIN 5.
FT   DOMAIN      870    885       SER-RICH.
FT   CARBOHYD    280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    409    409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    554    554       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    699    699       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     652    652       P -> T (IN ONE FORM).
SQ   SEQUENCE   913 AA;  100885 MW;  BD2BB9ECC815DDD6 CRC64;

  Query Match             16.5%;  Score 715;  DB 1;  Length 913;
  Best Local Similarity   26.9%;  Pred. No. 4.7e-36;
  Matches  203;  Conservative  117;  Mismatches  329;  Indels  106;  Gaps   19;

Qy    123 RPTFLQSKYEGSVRQNSRPGKPFLYV------------NATDLDDPATPNGQLYYQIVI 169
          :| | :  |   | ||   |:  | |                 ||     :| :|   |
Db     31 KPGFSEEDYTAFVSQNIMEGQKLLKVKFNNCAGNKGVRYETNSLDFKVRADGTMY--AVH 88

Qy    170 QLPMINNVMYFQI-------------------------------NNKTGAISLTREGSQE 198
          |: | :  :  :                                 |:| : | :: |
Db     89 QVQMASKQLILMVTAWDPQTLGRWEAIVRFLVGEKLQHNGHKPKGRKSGPVDLAQQQSDT 148

Qy    199 LNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIK 258
          | | :        | | :  |  :                 |  | |: : |||  | | :
Db    149 LLPWRQHQ-----SAKGLRRQKRD---------------WVIP-PINVPENSRGPFPQQ 186

Qy    259 ITQVRWN---DPGAQYSL--VDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVA 312
          : ::| :   :   :||:  |  :: |   ||||   | :|||:|:||||: :|   | |
Db    187 LVRIRSDKDKEIHIRYSITGVGADQPPMEVFSIDPVSGRMYVTRPMDREERASYHLRAHA 246

Qy    313 KDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENT 372
          |  |  :  |:::::| | |:||| |   : |   | |  : |  : |:||:| |:  |
Db    247 VDMNGNKVENPIDLYIYVIDMNDNRPEFINQVYNGSVDEGSKPGTYVMTVTANDADDSTT 306

Qy    373 ANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD---- 428
          ||  : |||| |||: |   :| | :  | :       | ::   || : :: :| :
Db    307 ANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGLDREKVQQYMVIVQATDMEGNLN 366

Qy    429 --FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
                | | |:||  | |  | |       :  :   :  :   | |:| : :
Db    367 YGLSNTATAIITVTDVNDNPPEFTTSTYSG--EVPENRVEVVVANLTVMDRDQPHSPNWN 424

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
          :| || ||  |   : ||| || | | : | :|:|      :      |  ||  |:: :
Db    425 AIYRIISGDPSGHFTIRTDPVTNEGMVTVVKAVDYEMNRAFMLTVMVSNQAPLASGIQMS 484

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLR 603
            | |  |: |||||||| |  :    :: | |  || :   :| ||:      : ||
Db    485 FQSTAGVTISVTDVNEAPYFPTNHKLIRLEEGVPTGTVLTTFSAVDPDRFMQQAVRYSKL 544
```

```
Qy    604 GDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMD 659
          |   || |:   |:| : | ||||:     : |    :| : |    |      : |:|
Db    545 SDPANWLNINATNGQITTAAVLDRESDYIKNNVYEATFLAADNGIPPASGTGTLQIYLID 604

Qy    660 VNDNPPRLAKDYTGLFFCHPLSAPGSLIFEAT----DDDQHLFRGPHFTFSLGS--GSLQ 713
          :||| | |      :       |   :   |    | | ::  || | | | |   :::
Db    605 INDNAPELLPKEAQI-----CEKPNLNVINITAADADIDPNV--GP-FVFELPSVPSAVR 656

Qy    714 NDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCF 773
           :| ::::|| :|:|| |    |   | | | : | | |||     : |  | | |
Db    657 KNWTITRLNGDYAQLSLRIMYLEAGVYDVPIIVTDSGNPPLYNTSIIKVKVCPCDENGDC 716

Qy    774 RPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFI 808
            |    :   |:  | ::  |: | |:| :| :
Db    717 TTIG---AVAAAGLGTGAIIAILICIIILLTMVLL 748


RESULT 8
CAD4_MOUSE
ID   CAD4_MOUSE     STANDARD;      PRT;   913 AA.
AC   P39038;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
GN   CDH4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=94067164; PubMed=8247017;
RA   Hutton J.C., Christofori G., Chi W.Y., Edman U., Guest P.C.,
RA   Hanahan D., Kelly R.B.;
RT   "Molecular cloning of mouse pancreatic islet R-cadherin: differential
RT   expression in endocrine and exocrine tissue.";
RL   Mol. Endocrinol. 7:1151-1160(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94095672; PubMed=8270638;
RA   Matsunami H., Miyatani S., Inoue T., Copeland N.G., Gilbert D.,
RA   Jenkins N.A., Takeichi M.;
RT   "Cell binding specificity of mouse R-cadherin and chromosomal mapping
RT   of the gene.";
RL   J. Cell Sci. 106:401-409(1993).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. MAY PLAY AN IMPORTANT ROLE
CC       IN RETINAL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: DISTRIBUTED WIDELY IN MOUSE TISSUES WITH HIGH
CC       LEVELS PRESENT IN BRAIN, SKELETAL MUSCLE, AND THYMUS.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X69966; CAA49589.1; -.
DR   EMBL; D14888; BAA03605.1; -.
DR   PIR; A47543; A47543.
DR   HSSP; P15116; 1NCI.
DR   MGD; MGI:99218; Cdh4.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21    166       POTENTIAL.
FT   CHAIN       167    913       CADHERIN-4.
FT   DOMAIN      167    731       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    732    753       POTENTIAL.
FT   DOMAIN      754    913       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      167    274       CADHERIN 1.
FT   DOMAIN      275    389       CADHERIN 2.
FT   DOMAIN      390    504       CADHERIN 3.
FT   DOMAIN      505    610       CADHERIN 4.
FT   DOMAIN      611    721       CADHERIN 5.
FT   DOMAIN      870    885       SER-RICH.
FT   CARBOHYD    146    146       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    409    409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    554    554       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    658    658       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    699    699       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   913 AA;  100030 MW;  1245A8CE8C338481 CRC64;

  Query Match             16.3%;  Score 705;  DB 1;  Length 913;
  Best Local Similarity   30.1%;  Pred. No. 1.9e-35;
  Matches  178;  Conservative  101;  Mismatches  280;  Indels   33;  Gaps   12;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
          |  | |: : |||  | | :: ::|    ||   :||:  |  :: |   |:||   | :
Db    168 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFNIDSMSGRM 226

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
          |||:|:|||||: :|   | | |  |  :  |:::::  | |:||| |   : |   | |
Db    227 YVTRPMDREERASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNRPEFINQVYNGSVDEG 286

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
          : |  : |:||:| |:  |||  : |||| |||: |   :| | :  | :      | :
Db    287 SKPGTYVMTVTANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGLDR 346

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
          :   || : :: :| :              | | |:||  | |  | :       : |
Db    347 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPPEFTTSTFAG--EVPENRIE 404

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
          : :  :   | |:| : :   :| || ||  |   | |||  || | | : | :|:|
Db    405 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 464

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
          :     |  ||  |:: :  | |  |: |||||||| |  :    :: | |  || :
Db    465 FMLTVMVSNQAPLASGIQMSFQSTAGVTISVTDVNEAPYFPSNHKLIRLEEGVPAGTALT 524

Qy    587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
          :| ||:      : ||   |   || |:   |:| : | ||||:    :  |
Db    525 TFSAVDPDRFMQQAVRYSKLSDPANWLHINTSNGQITTAAILDRESLYTKNNVYEATFLA 584

Qy    640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPG--SLIFEATDDDQHL 697
          : |    |      : |:|:||| |:|      :        ||  ::   | | |
Db    585 ADNGIPPASGTGTLQIYLIDINDNAPQLLPKEAQI-----CERPGLNAINITAADADMDP 639

Qy    698 FRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEG 756
          ||: |       ::: :| ::::|| :|:|| |    |   | | | : | | |||
Db    640 NIGPYVFELPFIPTTVRKNWTITRLNGDYAQLSLRILYLEAGVYDVPIIVTDSGNPPLSN 699

Qy    757 IVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFI 808
          : |  | | |       |    :   |:  | ::  |: | |:| :| :
Db    700 TSVIKVKVCPCDENGDCTTVG---AVAAAGLGTGAIVAILICIVILLIMVLL 748
```

```
RESULT 9
CAD2_HUMAN
ID   CAD2_HUMAN     STANDARD;      PRT;   906 AA.
AC   P19022; Q14923;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neural-cadherin precursor (N-cadherin) (Cadherin-2).
GN   CDH2 OR CDHN OR NCAD.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91016946; PubMed=2216790;
RA   Reid R.A., Hemperly J.J.;
RT   "Human N-cadherin: nucleotide and deduced amino acid sequence.";
RL   Nucleic Acids Res. 18:5896-5896(1990).
RN   [2]
RP   REVISIONS TO 341; 699 AND 705.
RA   Reid R.A.;
RL   Submitted (NOV-1990) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92363956; PubMed=1500442;
RA   Salomon D., Ayalon O., Patel-King R., Hynes R.O., Geiger B.;
RT   "Extrajunctional distribution of N-cadherin in cultured human
RT   endothelial cells.";
RL   J. Cell Sci. 102:7-17(1992).
RN   [4]
RP   SEQUENCE OF 160-906 FROM N.A.
RX   MEDLINE=90347462; PubMed=2384753;
RA   Walsh F.S., Barton C.H., Putt W., Moore S.E., Kelsell D.,
RA   Spurr N., Goodfellow P.N.;
RT   "N-cadherin gene maps to human chromosome 18 and is not linked to the
RT   E-cadherin gene.";
RL   J. Neurochem. 55:805-812(1990).
RN   [5]
RP   SEQUENCE OF 1-20 FROM N.A.
RX   MEDLINE=95048366; PubMed=7959764;
RA   Wallis J.A., Fox M., Walsh F.S.;
RT   "Structure of the human N-cadherin gene: YAC analysis and fine
RT   chromosomal mapping to 18q11.2.";
RL   Genomics 22:172-179(1994).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X57548; CAA40773.1; -.
DR   EMBL; X54315; CAA38213.1; -.
DR   EMBL; S42303; AAB22854.1; -.
DR   EMBL; M34064; AAA03236.1; -.
DR   EMBL; Z27420; CAA81799.1; -.
DR   PIR; A38870; IJHUCN.
DR   HSSP; P15116; 1NCJ.
DR   Genew; HGNC:1759; CDH2.
DR   MIM; 114020; -.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     23       POTENTIAL.
FT   PROPEP       24    159
FT   CHAIN       160    906       NEURAL-CADHERIN.
FT   DOMAIN      160    724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    725    746       POTENTIAL.
FT   DOMAIN      747    906       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      160    267       CADHERIN 1.
FT   DOMAIN      268    382       CADHERIN 2.
FT   DOMAIN      383    497       CADHERIN 3.
FT   DOMAIN      498    603       CADHERIN 4.
FT   DOMAIN      604    714       CADHERIN 5.
FT   DOMAIN      863    878       SER-RICH.
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    273    273       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    572    572       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    692    692       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     12     12       MISSING (IN REF. 3).
FT   CONFLICT     16     16       L -> A (IN REF. 3 AND 5).
FT   CONFLICT    196    196       S -> T (IN REF. 1).
FT   CONFLICT    212    212       I -> L (IN REF. 4).
FT   CONFLICT    357    357       N -> I (IN REF. 1).
FT   CONFLICT    867    867       A -> L (IN REF. 3).
SQ   SEQUENCE   906 AA;  99851 MW;  72DDC7B8B57C7AFC CRC64;

  Query Match             16.2%;  Score 699.5;  DB 1;  Length 906;
  Best Local Similarity   27.4%;  Pred. No. 4.1e-35;
  Matches  210;  Conservative  122;  Mismatches  331;  Indels  103;  Gaps   21;

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
          ||::: :|  |  |:      |    :   | ::   |     :| :|   |   |: :
Db     47 KDVHEGQP-LLNVKF-----SNCNGKRKVQYESSEPADFKVDEDGMVY--AVRSFPLSSE 98

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234
          | |              :|:    :    : :|:|     :| |  ::  |: ||
Db     99 HAKFLIY-----------AQDKETQEKWQVAVKLSLKPT--LTEESVKESAEVEEIVFP 144

Qy    235 ------------TENIWKAPKPVEMVENSTDPHPIKITQVRWN------------DPGAQ 270
                      :  |  | |: : |||  | | :: ::| :             |||
Db    145 RQFSKHSGHLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA- 202

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329
            :: |   | |:   | : ||:|||||:   :   | | |  |  :  |::| :
Db    203 ------DQPPTGIFIINPISGQLSVTKPLDREQIARFHLRAHAVDINGNQVENPIDIVIN 256

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389
          | |:||| |     |    | | : | : |:|| | |: |  |  | |||| | |  |
Db    257 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRIVSQAPSTP 316

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443
             :| |    | :      | ::   || | |: :| :                | | |:|
Db    317 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVN 376

Qy    444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502
          |  | |   : || :   :  :   :  || |:| | :   :| |  ||  ||  :
Db    377 DNPPEFTAMTFYGEV---PENRVDIIVANLTVTDKDQPHTPAWNAVYRISGGDPTGRFAI 433

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562
          |||::| | | : ||:|||| :  :   ||| ||  |:::   | |  :: | ||||
Db    434 QTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 493

Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
            | |: :    :  | :  || :   ||:||:     :| |:   |   ||||| | |:|
```

```
Db    494 NPYFAPNPKIIRQEEGLHAGTMLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQI 553

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675
          ::| ||||:     : |    :|:: |   :|      : |:|:||| |::
Db    554 TTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL------- 606

Qy    676 FCHPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHAR 727
             | |        | |:   | | |     ||  |   |   ::: :| ::::|| |:
Db    607 ---PQEAETCETPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQ 663

Qy    728 LSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV- 785
          |: :    |   | | | | | | ||   |  | |  | |   | |       | :  :
Db    664 LNLKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIV 716

Qy    786 --GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
            |:  | ::  || | |:| :| : :   | :|    |:   : |
Db    717 GAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762


RESULT 10
CAD2_CHICK
ID   CAD2_CHICK     STANDARD;      PRT;   912 AA.
AC   P10288; Q90630;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-MAR-1989 (Rel. 10, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neural-cadherin precursor (N-cadherin) (Cadherin-2).
GN   CDH2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=88153917; PubMed=2831236;
RA   Hatta K., Nose A., Nagafuchi A., Takeichi M.;
RT   "Cloning and expression of cDNA encoding a neural calcium-dependent
RT   cell adhesion molecule: its identity in the cadherin gene family.";
RL   J. Cell Biol. 106:873-881(1988).
RN   [2]
RP   SEQUENCE OF 1-25 FROM N.A.
RC   STRAIN=Cornish white rock Cockerel;
RX   MEDLINE=97354288; PubMed=9210582;
RA   Li B., Paradies N.E., Brackenbury R.W.;
RT   "Isolation and characterization of the promoter region of the chicken
RT   N-cadherin gene.";
RL   Gene 191:7-13(1997).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X07277; CAA30258.1; -.
DR   EMBL; U15563; AAB62980.1; -.
DR   PIR; A29964; IJCHCN.
DR   HSSP; P15116; 1NCJ.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     28       POTENTIAL.
FT   PROPEP       29    164       POTENTIAL.
FT   CHAIN       165    912       NEURAL-CADHERIN.
FT   DOMAIN      165    729       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    730    752       POTENTIAL.
FT   DOMAIN      753    912       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      165    272       CADHERIN 1.
FT   DOMAIN      273    387       CADHERIN 2.
FT   DOMAIN      388    502       CADHERIN 3.
FT   DOMAIN      503    609       CADHERIN 4.
FT   DOMAIN      610    720       CADHERIN 5.
FT   DOMAIN      869    884       SER-RICH.
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    407    407       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    578    578       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    628    628       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    657    657       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     21     21       A -> G (IN REF. 2).
SQ   SEQUENCE   912 AA;  100464 MW;  9BA5AC9DC1FFC489 CRC64;

  Query Match             16.1%;  Score 696;  DB 1;  Length 912;
  Best Local Similarity   26.9%;  Pred. No. 6.7e-35;
  Matches  209;  Conservative  117;  Mismatches  328;  Indels  124;  Gaps   22;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQIN-- 183
          | :  :   | ::   |:| | |     |:                     |  :|| :
Db     41 FPEDVHSAVVSRSVHGGQPLLNVRFQSCDE------------------NRKIYFGSSEP 81

Qy    184 -----NKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQS----------ENSFSDTT 228
               : | :   |       | :      |:| :|   |           | :|:  :
Db     82 EDFRVGEDGVVYAERSFQLSAEPTE-----FVVSARDKETQEEWQMKVKLTPEPAFTGAS 136

Qy    229 SVDIIVTENI------------------WKAPKPVEMVENSTDPHPIKITQVRWNDPGA 269
            |    |:|                  |  | |: : |||  | | :: ::|
Db    137 EKDQKKIEDIIFPWQQYKDSSHLKRQKRDWVIP-PINLPENSRGPFPQELVRIR------ 189

Qy    270 QYSLVDKEKLPRFPFSI-----DQ----------EGDIYVTQPLDREEKDAYVFYAVAK 313
               |::|     :|:     ||           | : ||:|||||:  ::   | |
Db    190 ----SDRDKSLSLRYSVTGPGADQPPTGIFIINPISGQLSVTKPLDREQIASFHLRAHAV 245

Qy    314 DEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTA 373
          |  |  :  |::| : | |:||| |     |    | |  : |  : |:|| | |: |
Db    246 DVNGNQVENPIDIVINVIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNAQ 305

Qy    374 NSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD----- 428
          |  | |||: | |  |   :| |    | :    | ::   || | |: :| :
Db    306 NGMLRYRILSQAPSSPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTY 365

Qy    429 -FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSK 486
               | | |:||  | |   : || :    :  :   :  :  || |:| | :
Db    366 GLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWN 422

Qy    487 ILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYN 546
          | :  ||  |:  : |||::| | | : ||:||||  :  :   |||  ||  |:::
Db    423 ARYQMTGGDPTGQFTILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHP 482

Qy    547 ASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDP----EGLDISYSL 602
          | |  :: | ||||:| | :    :  | :  |: :   ||:||    :   : ||
Db    483 PQSTATVSITVIDVNESPYFVPNPKLVRQEEGLLAGSMLTTFTARDPDRYMQQTSLRYSK 542

Qy    603 RGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILM 658
          |   ||||| | |:| : | ||||:     : |    :|:: |   :|      : |:
Db    543 LSDPANWLKIDPVNGQITTTAVLDRESIYVQNNMYNATFLASDNGIPPMSGTGTLQIYLL 602

Qy    659 DVNDNPPRL-AKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSG--SLQND 715
          |:||| |::  |: |    |  |  | ::   | | |     || | | |     |:: :
Db    603 DINDNAPQVNPKEAT---TCETLQ-PNAINITAVDPDIDPNAGP-FAFELPDSPPSIKRN 657
```

```
Qy    716 WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFR 774
          | : :|:| ||:|| |    |   | | | | | | |       | |  | | : | |
Db    658 WTIVRISGDHAQLSLRIRFLEAGIYDVPIVITDSGNPHASSTSVLKVKVCQCDINGDC-- 715

Qy    775 PAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
                | :  :   |:  | ::  || | |:| :| : :   | :|    |:   : |
Db    716 -----TDVDRIVGAGLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 768


RESULT 11
CAD2_MOUSE
ID   CAD2_MOUSE     STANDARD;      PRT;   906 AA.
AC   P15116; Q64260;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neural-cadherin precursor (N-cadherin) (Cadherin-2).
GN   CDH2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89346748; PubMed=2762814;
RA   Miyatani S., Shimamura K., Hatta M., Nagafuchi A., Nose A.,
RA   Matsunaga M., Hatta K., Takeichi M.;
RT   "Neural cadherin: role in selective cell-cell adhesion.";
RL   Science 245:631-635(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Tamura K.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6;
RX   MEDLINE=92409532; PubMed=1528849;
RA   Miyatani S., Copeland N.G., Gilbert D.J., Jenkins N.A., Takeichi M.;
RT   "Genomic structure and chromosomal mapping of the mouse N-cadherin
RT   gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:8443-8447(1992).
RN   [4]
RP   DEVELOPMENTAL STAGE.
RC   STRAIN=C57BL/6; TISSUE=Testis;
RX   MEDLINE=97033837; PubMed=8879495;
RA   Munro S.B., Blaschuk O.W.;
RT   "A comprehensive survey of the cadherins expressed in the testes of
RT   fetal, immature, and adult mice utilizing the polymerase chain
RT   reaction.";
RL   Biol. Reprod. 55:822-827(1996).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 160-267.
RX   MEDLINE=95191680; PubMed=7885471;
RA   Shapiro L., Fannon A.M., Kwong P.D., Thompson A., Lehmann M.S.,
RA   Gruebel G., Legrand J.-F., Als-Nielsen J., Colman D.R.,
RA   Hendrickson W.A.;
RT   "Structural basis of cell-cell adhesion by cadherins.";
RL   Nature 374:327-337(1995).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (3.4 ANGSTROMS) OF 160-374.
RX   MEDLINE=98318235; PubMed=9655503;
RA   Tamura K., Shan W.S., Hendrickson W.A., Colman D.R., Shapiro L.;
RT   "Structure-function analysis of cell adhesion by neural (N-)
RT   cadherin.";
RL   Neuron 20:1153-1163(1998).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT ALL STAGES OF TESTICULAR
CC       DEVELOPMENT WITH HIGHEST LEVELS FOUND IN TESTES OF 21-DAY-OLD
CC       MICE.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M31131; AAA37353.1; -.
DR   EMBL; AB008811; BAA23549.1; -.
DR   EMBL; S45011; AAB23356.1; -.
DR   PIR; A32759; IJMSCN.
DR   PDB; 1NCG; 10-JUL-95.
DR   PDB; 1NCH; 10-JUL-95.
DR   PDB; 1NCI; 10-JUL-95.
DR   PDB; 1NCJ; 18-MAR-99.
DR   MGD; MGI:88355; Cdh2.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal; 3D-structure.
FT   SIGNAL       1     23       POTENTIAL.
FT   PROPEP      24    159       POTENTIAL.
FT   CHAIN      160    906       NEURAL-CADHERIN.
FT   DOMAIN     160    724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   725    746       POTENTIAL.
FT   DOMAIN     747    906       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     160    267       CADHERIN 1.
FT   DOMAIN     268    382       CADHERIN 2.
FT   DOMAIN     383    497       CADHERIN 3.
FT   DOMAIN     498    603       CADHERIN 4.
FT   DOMAIN     604    717       CADHERIN 5.
FT   DOMAIN     863    878       SER-RICH.
FT   CARBOHYD   190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   273    273       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   572    572       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   692    692       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT     7      9       GRG -> APR (IN REF. 2).
FT   CONFLICT   565    565       Y -> N (IN REF. 2).
FT   CONFLICT   567    567       Q -> K (IN REF. 2).
FT   CONFLICT   624    624       A -> T (IN REF. 2).
SQ   SEQUENCE   906 AA;  99761 MW;  0206741B771707B7 CRC64;

  Query Match           16.1%;  Score 694.5;  DB 1;  Length 906;
  Best Local Similarity   27.6%;  Pred. No. 8.2e-35;
  Matches  211;  Conservative  122;  Mismatches  332;  Indels   99;  Gaps   23;

Qy    117 KDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINN 176
          ||::: :|  |  |:    |:     :  | ::   |     :| :|  |   |:
Db     47 KDVHEGQP-LLNVKFSNCNRK-----RKVQYESSEPADFKVDEDGTVY--AVRSFPLTAE 98

Qy    177 VMYFQI---NNKTG-----AISLTREGSQELNPAKNP-SYNLVISVKDM---GGQSENSF 224
             | |   : :|      |::|:|| :    | | |     ::  : :    |  :
Db     99 QAKFLIYAQDKETQEKWQVAVNLSREPTLTEEPMKEPHEIEEIVFPRQLAKHSGALQRQK 158

Qy    225 SDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIKITQVRWN-----------DPGAQYS 272
            .|           | | |: : |||  | | :: ::| :            |||
Db    159 RD-----------WVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA--- 202

Qy    273 LVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVK 331
```

```
Db    203 ----DQPPTGIFIINPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVINVI 258

Qy    332 DINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMD 391

Db    259 DMNDNRPEFLHQVWNGSVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTPSP 318

Qy    392 GLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQ 445

Db    319 NMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVNDN 378

Qy    446 IPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDT 504

Db    379 PPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWNAAYRISGGDPTGRFAILT 435

Qy    505 DPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAP 564

Db    436 DPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNENP 495

Qy    565 QFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFS 621

Db    496 YFAPNPKIIRQEEGLHAGTMLTTLTAQDPDRYMQQNIRYTKLSDPANWLKIDPVNGQITT 555

Qy    622 VAPLDREA----GSPYRVQVVATEVGGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFC 677

Db    556 IAVLDRESPYVQNNIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVL--------- 606

Qy    678 HPLSA-------PGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLS 729

Db    607 -PQEAETCETPEPNSINIAALDYDIDPNAGPFAFDLPLSPVTIKRNWTINRLNGDFAQLN 665

Qy    730 TRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV--- 785

Db    666 LKIKFLEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIVGA 718

Qy    786 GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829

Db    719 GLGTGAIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 762


RESULT 12
CAD2_BOVIN
ID   CAD2_BOVIN     STANDARD;      PRT;   877 AA.
AC   P19534;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neural-cadherin precursor (N-cadherin) (Cadherin-2) (Fragment).
GN   CDH2.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90360979; PubMed=2390969;
RA   Liaw C.W., Cannon C., Power M.D., Kiboneka P.K., Rubin L.L.;
RT   "Identification and cloning of two species of cadherins in bovine
RT   endothelial cells.";
RL   EMBO J. 9:2701-2708(1990).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X53615; CAA37677.1; -.
DR   PIR; S11693; IJBOCN.
DR   HSSP; P15116; 1NCI.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT   NON_TER       1      1
FT   PROPEP       <1    130
FT   CHAIN       131    877       NEURAL-CADHERIN.
FT   DOMAIN      131    695       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    696    717       POTENTIAL.
FT   DOMAIN      718    877       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      131    238       CADHERIN 1.
FT   DOMAIN      239    363       CADHERIN 2.
FT   DOMAIN      364    468       CADHERIN 3.
FT   DOMAIN      469    574       CADHERIN 4.
FT   DOMAIN      575    685       CADHERIN 5.
FT   DOMAIN      834    849       SER-RICH.
FT   CARBOHYD    161    161       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    244    244       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    296    296       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    373    373       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    543    543       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    622    622       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    663    663       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   877 AA;  96845 MW;  441B829ED871A249 CRC64;

  Query Match             15.9%;  Score 688.5;  DB 1;  Length 877;
  Best Local Similarity   26.7%;  Pred. No. 1.8e-34;
  Matches  203;  Conservative  126;  Mismatches  343;  Indels   87;  Gaps   19;

Qy    126 FLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLP---------MINN 176

Db      7 FPEDVYSAVLSRDVLEGQPLLNVKFSNC------NGKRKVQYESSEPADFKVDEDGMVYA 60

Qy    177 VMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIV-- 234

Db     61 VRSFPLSSEHSKFLIYAQDKE-----TQEKWQVAVKLSLKPALPEDSVKESREIEEIVFP 115

Qy    235 ------------TENIWKAPKPVEMVENSTDPHPIKITQVRWN------------DPGAQ 270

Db    116 RQVTKHNGYLQRQKRDWVIP-PINLPENSRGPFPQELVRIRSDRDKNLSLRYSVTGPGA- 173

Qy    271 YSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVK 329

Db    174 ------DQPPTGIFIINPISGQLSVTKPLDRELIARFHLRAHAVDINGNQVENPIDIVIN 227

Qy    330 VKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLP 389

Db    228 VIDMNDNRPEFLHQVWNGTVPEGSKPGTYVMTVTAIDADDPNALNGMLRYRILSQAPSTP 287

Qy    390 MDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDIN 443

Db    288 SPNMFTINNETGDIITVAAGLDREKVQQYTLIIQATDMEGNPTYGLSNTATAVITVTDVN 347

Qy    444 DQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGV 502

Db    348 DNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKDQPHTPAWNAIYRISGGDPAGRFAI 404

Qy    503 DTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE 562

Db    405 QTDPNSNDGLVTVVKPIDFETNRMYVLTVAAENQVPLAKGIQHPPQSTATVSVTVIDVNE 464
```

```
Qy    563 APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEI 619
          | |: :     :  | :  || :   ||:||:     :| |:   |   ||||| | |:|
Db    465 NPYFAPNPKIIRQEEGLHAGTVLTTFTAQDPDRYMQQNIRYTKLSDPANWLKIDSVNGQI 524

Qy    620 FSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLF 675
          ::| ||||:     : |    :|:: |   :|      : |:|:||| |::      :
Db    525 TTIAVLDRESPNVKANIYNATFLASDNGIPPMSGTGTLQIYLLDINDNAPQVLPQEAEI- 583

Qy    676 FCHPLSAPGSLIFEATDDDQHLFRGPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTD 734
          |     | |:   | | |     ||  |   |   ::: :| ::::||  |:|: :
Db    584 -CE-TPDPNSINITALDYDIDPNAGPFAFDLPLSPVTIKRNWTITRLNGDFAQLNLKIKF 641

Qy    735 FEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTV---GMAVG 790
          |   | | | | | | ||   |  | | | | |   | |       | :  :   |:  |
Db    642 LEAGIYEVPIIITDSGNPPKSNISILRVKVCQCDSNGDC-------TDVDRIVGAGLGTG 694

Qy    791 ILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
          ::  || | |:| :| : :   | :|    |:   : |
Db    695 AIIAILLCIIILLILVLMFVVWMKRRDKERQAKQLLIDP 733


RESULT 13
CAD2_RAT
ID   CAD2_RAT       STANDARD;      PRT;   906 AA.
AC   Q9Z1Y3; Q9R0T5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neural-cadherin precursor (N-cadherin) (Cadherin-2).
GN   CDH2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   STRAIN=Sprague-Dawley; TISSUE=Testis;
RX   MEDLINE=98187820; PubMed=9528971;
RA   Chung S.S., Mo M.Y., Silvestrini B., Lee W.M., Cheng C.Y.;
RT   "Rat testicular N-cadherin: its complementary deoxyribonucleic acid
RT   cloning and regulation.";
RL   Endocrinology 139:1853-1862(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague-Dawley; TISSUE=Brain;
RA   Asai K., Tada T., Yamamoto M., Tada A., Mizuno M., Eimoto T., Kato T.;
RL   Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC   -!- TISSUE SPECIFICITY: IN TESTIS, EXPRESSED IN SERTOLI AND GERM
CC       CELLS.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF097593; AAC83818.1; -.
DR   EMBL; AB017695; BAA84919.1; -.
DR   HSSP; P15116; 1NCJ.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     27       POTENTIAL.
FT   PROPEP       28    159       POTENTIAL.
FT   CHAIN       160    906       NEURAL-CADHERIN.
FT   DOMAIN      160    724       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    725    745       POTENTIAL.
FT   DOMAIN      746    906       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      160    267       CADHERIN 1.
FT   DOMAIN      268    382       CADHERIN 2.
FT   DOMAIN      383    497       CADHERIN 3.
FT   DOMAIN      498    603       CADHERIN 4.
FT   DOMAIN      604    717       CADHERIN 5.
FT   DOMAIN      863    878       SER-RICH.
FT   CARBOHYD    190    190       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    273    273       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    325    325       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    402    402       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    572    572       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    622    622       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    651    651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    692    692       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      7      7       G -> A (IN REF. 2).
FT   CONFLICT     48     48       T -> D (IN REF. 2).
FT   CONFLICT    153    153       A -> R (IN REF. 2).
FT   CONFLICT    646    646       A -> V (IN REF. 2).
FT   CONFLICT    658    658       N -> K (IN REF. 2).
FT   CONFLICT    724    724       T -> A (IN REF. 2).
SQ   SEQUENCE   906 AA;  99685 MW;  97D9937FD8D1F8B5 CRC64;

  Query Match           15.9%;  Score 686.5;  DB 1;  Length 906;
  Best Local Similarity  27.2%;  Pred. No. 2.5e-34;
  Matches  215;  Conservative  126;  Mismatches  344;  Indels  105;  Gaps   24;

Qy     97 AALDANGIIV------EGPVPITIEVKDINDNRPTFLQSKYEGSVRQNSRPGKPFLYVNA 150
          |:|:|:| :          |   :  | ::: :|  |  |:  . |:     :  | ::
Db     21 ASLEASGELALCKTGFPEDVYSAVLPKTVHEGQP-LLNVKFSNCNRK-----RKVQYESS 74

Qy    151 TDLDDPATPNGQLYYQIVIQLPMINNVMYFQI---NNKTG-----AISLTREGSQELNPA 202
             |     :| :|   |   |:       | |   : :|      |::|: | |    |
Db     75 EPADFKVDEDGTVY--AVRSFPLSAEQAKFLIYAQDKETQEKWQVAVNLSLEPSLTEEPM 132

Qy    203 KNP-SYNLVISVKDM---GGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIK 258
          | |      ::  : :    |  :    |            |  | |: : |||  | | :
Db    133 KEPHEIEEIVFPRQLAKHSGALQRQKRD-----------WVIP-PINLPENSRGPFPQE 179

Qy    259 ITQVRWN-----------DPGAQYSLVDKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDA 305
          : ::| :             |||       :: :|   | |:   | : ||:|||||
Db    180 LVRIRSDRDKNLSLRYSVTGPGA-------DQPPTGIFIINPISGQLSVTKPLDRELIAR 232

Qy    306 YVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAH 365
          :  | | |  |  :  |::| : | |:||| |     |    | |  : |  : |:||
Db    233 FHLRAHAVDINGNQVENPIDIVINVIDMNDNRPEFLHQVWNGSVPEGSKPGTYVMTVTAI 292

Qy    366 DRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVS 425
          | |: |  |  | |||: | |  |   :| |   | :      | ::   || | |: :
Db    293 DADDPNALNGMLRYRILSQAPSTPSPNMFTINNETGDIITVAAGLDREKVQQYTLIIQAT 352

Qy    426 DKD------FKTLCFVQINVIDINDQIPIF-EKSDYGNLTLAEDTNIGSTILTIQATDAD 478
          | :             | | |:||  | |   : || :    :  :   :  : || |
Db    353 DMEGNPTYGLSNTATAVITVTDVNDNPPEFTAMTFYGEV---PENRVDVIVANLTVTDKD 409

Qy    479 EPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEP 538
          :| | :    | |  ||  ||  : |||::| | | : ||:||||  :  :   |||  |
Db    410 QPHTPAWNAAYRISGGDPTGRFAILTDPNSNDGLVTVVKPIDFETNRMFVLTVAAENQVP 469

Qy    539 LVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL-- 596
          |  |:::   | |  :: | |||| | |: :    :  | :  || :  :||:||:
Db    470 LAKGIQHPPQSTATVSVTVIDVNENPYFAPNPKIIRQEEGLHAGTMLTTLTAQDPDRYMQ 529
```

```
Qy    597 -DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVS 651
          :| |:   |   ||||| | |:| ::| ||||:     : |    :|:: |   :|
Db    530 QNIRYTKLSDPANWLKIDPVNGQITTIAVLDRESPNVKNNIYNATFLASDNGIPPMSGTG 589

Qy    652 EFHLILMDVNDNPPRLAKDYTGLFFCHPLSA-------PGSLIFEATDDDQHLFRGPH-F 703
            : |:|:||| |::           |  |       | |:   | | |     ||  |
Db    590 TLQIYLLDINDNAPQVL----------PQEAETCETPEPNSINITALDYDIDPNAGPFAF 639

Qy    704 TFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVT 763
            |   ::: :| ::::|| |:|: :    |   | | | | | | ||   |  | |
Db    640 DLPLSPATIKRNWTITRLNGDFAQLNLKIKFLEAGIYEVPIVITDSGNPPKSNISILRVK 699

Qy    764 FCSC-VEGSCFRPAGHQTGIPTV---GMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNV 819
          | |   | |       | :  :   |:  | ::  || | |:| :| : :   | :|
Db    700 VCQCDSNGDC-------TDVDRIVGAGLGTGTIIAILLCIIILLILVLMFVVWMKRRDKE 752

Qy    820 ESAQASEVKP 829
            |:   : |
Db    753 RQAKQLLIDP 762


RESULT 14
CAD4_HUMAN
ID   CAD4_HUMAN     STANDARD;      PRT;   916 AA.
AC   P55283;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Cadherin-4 precursor (Retinal-cadherin) (R-cadherin) (R-CAD).
GN   CDH4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95073006; PubMed=7982033;
RA   Tanihara H., Sano K., Heimark R.L., St John T., Suzuki S.;
RT   "Cloning of five human cadherins clarifies characteristic features of
RT   cadherin extracellular domain and provides further evidence for two
RT   structurally different types of cadherin.";
RL   Cell Adhes. Commun. 2:15-26(1994).
RN   [2]
RP   SEQUENCE OF 393-916 FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=91283540; PubMed=2059658;
RA   Suzuki S., Sano K., Tanihara H.;
RT   "Diversity of the cadherin family: evidence for eight new cadherins
RT   in nervous tissue.";
RL   Cell Regul. 2:261-270(1991).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. MAY PLAY AN IMPORTANT ROLE
CC       IN RETINAL DEVELOPMENT.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED MAINLY IN BRAIN BUT ALSO FOUND IN
CC       OTHER TISSUES.
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L34059; AAA35627.1; -.
DR   HSSP; P15116; 1NCI.
DR   Genew; HGNC:1763; CDH4.
DR   MIM; 603006; -.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW   Signal.
FT   SIGNAL        1     20       POTENTIAL.
FT   PROPEP       21    169       POTENTIAL.
FT   CHAIN       170    916       CADHERIN-4.
FT   DOMAIN      170    734       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    735    756       POTENTIAL.
FT   DOMAIN      757    916       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      170    277       CADHERIN 1.
FT   DOMAIN      278    392       CADHERIN 2.
FT   DOMAIN      393    507       CADHERIN 3.
FT   DOMAIN      508    613       CADHERIN 4.
FT   DOMAIN      614    724       CADHERIN 5.
FT   DOMAIN      873    888       SER-RICH.
FT   CARBOHYD    283    283       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    412    412       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    557    557       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    632    632       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    661    661       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    702    702       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   916 AA;  100446 MW;  E717C54A19E0C52A CRC64;

  Query Match             15.8%;  Score 684.5;  DB 1;  Length 916;
  Best Local Similarity   29.3%;  Pred. No. 3.4e-34;
  Matches  176;  Conservative  106;  Mismatches  288;  Indels   31;  Gaps   12;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRW---NDPGAQYSL--VDKEKLPRFPFSID-QEGDI 292
          |  | |: : |||  | | :: ::|    ||   :||:  |  :: |   |||:   | :
Db    171 WVIP-PINVPENSRGPFPQQLVRIRSDKDNDIPIRYSITGVGADQPPMEVFSINSMSGRM 229

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
          |||:|:|||| :|   | | |  |  :  |:::: | |:||| |   : |    | |
Db    230 YVTRPMDREEHASYHLRAHAVDMNGNKVENPIDLYIYVIDMNDNHPEFINQVYNCSVDEG 289

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
          : |  : |:||:| |:  |||  : |||| |||: |   :| | :  | :       :
Db    290 SKPGTYVMTITANDADDSTTANGMVRYRIVTQTPQSPSQNMFTINSETGDIVTVAAGWDR 349

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
          :   || : :: :| :              | | |:||    |  | :       : ::
Db    350 EKVQQYTVIVQATDMEGNLNYGLSNTATAIITVTDVNDNPSEFTASTFAG--EVPENSVE 407

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
          : :  :   | |:| : :   :| || ||  |   | ||| || | | : | :|:|
Db    408 TVVANLTVMDRDQPHSPNWNAVYRIISGDPSGHFSVRTDPVTNEGMVTVVKAVDYELNRA 467

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
            :     |  ||  |:: :  | |  |: : |:|||| |  :    :: | |  || :
Db    468 FMLTVMVSNQAPLASGIQMSFQSTAGVTISIMDINEAPYFPSNHKLIRLEEGVPPGTVLT 527

Qy    587 NVTAKDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVA 639
            :| ||:      : ||   |   || |:  |:| :|| ||||:     : |    :|
Db    528 TFSAVDPDRFMQQAVRYSKLSDPASWLHINATNGQITTVAVLDRESLYTKNNVYEATFLA 587

Qy    640 TEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFR 699
          : |    |      : |:|:||| | |      :    |:|   :  | | | |
Db    588 ADNGIPPASGTGTLQIYLIDINDNAPELLPKEAQICERPNLNA---INITAADADVHPNI 644

Qy    700 GPH-FTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIV 758
          ||: |      ::: :| ::::|| :|:|| |    |   | | | : | | |||
Db    645 GPYVFELPFVPAAVRKNWTITRLNGDYAQLSLRILYLEAGMYDVPIIVTDSGNPPLSNTS 704

Qy    759 SLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--FIRIKKDKGK 816
          : |  | | :       |    :   |:  | ::  |: | |:| :|  |:   | : |
Db    705 IIKVKVCPCDDNGDCTTIG---AVAAAGLGTGAIVAILICILILLTMVLLFVMWMKRREK 761
```

```
Qy    817 D 817
          :
Db    762 E 762


RESULT 15
DSC2_MOUSE
ID   DSC2_MOUSE     STANDARD;      PRT;   902 AA.
AC   P55292; Q64734;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Desmocollin 2A/2B precursor (Epithelial type 2 desmocollin).
GN   DSC2 OR DSC3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Embryo;
RX   MEDLINE=95227276; PubMed=7711832;
RA   Lorimer J.E., Hall L.S., Clarke J.P., Collins J.E., Fleming T.P.,
RA   Garrod D.R.;
RT   "Cloning, sequence analysis and expression pattern of mouse
RT   desmocollin 2 (DSC2), a cadherin-like adhesion molecule.";
RL   Mol. Membr. Biol. 11:229-236(1994).
RN   [2]
RP   SEQUENCE OF 344-637 FROM N.A.
RC   STRAIN=C57BL/6; TISSUE=Embryo;
RX   MEDLINE=95048328; PubMed=7959727;
RA   Buxton R.S., Wheeler G.N., Pidsley S.C., Marsden M.D., Adams M.J.,
RA   Jenkins N.A., Gilbert D.J., Copeland N.G.;
RT   "Mouse desmocollin (Dsc3) and desmoglein (Dsg1) genes are closely
RT   linked in the proximal region of chromosome 18.";
RL   Genomics 21:510-516(1994).
CC   -!- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED
CC       IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS
CC       MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL
CC       POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL
CC       ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 2A (SHOWN HERE) AND 2B; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: IN ALL EPITHELIA TESTED AND HEART.
CC   -!- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC       (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L33779; AAA79177.1; -.
DR   EMBL; L33779; AAA79176.1; -.
DR   EMBL; X73885; CAA52089.1; -.
DR   HSSP; P15116; 1NCJ.
DR   MGD; MGI:103221; Dsc2.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 3.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Alternative splicing; Cytoskeleton; Calcium-binding.
FT   SIGNAL        1     27       POTENTIAL.
FT   PROPEP       28    135       POTENTIAL.
FT   CHAIN       136    902       DESMOCOLLIN 2A/2B.
FT   DOMAIN      136    694       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    695    715       POTENTIAL.
FT   DOMAIN      716    902       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      136    243       CADHERIN 1.
FT   DOMAIN      244    355       CADHERIN 2.
FT   DOMAIN      356    471       CADHERIN 3.
FT   DOMAIN      472    579       CADHERIN 4.
FT   DOMAIN      580    694       CADHERIN 5.
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    392    392       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    546    546       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    629    629       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    838    848       KVQFCHTDDNQ -> ETIRGHTLIKN (IN ISOFORM
FT                                2B).
FT   VARSPLIC    849    902       MISSING (IN ISOFORM 2B).
SQ   SEQUENCE   902 AA;  99961 MW;  E233F8CB20ACCAB0 CRC64;

  Query Match             15.6%;  Score 674.5;  DB 1;  Length 902;
  Best Local Similarity   31.2%;  Pred. No. 1.4e-33;
  Matches  185;  Conservative   99;  Mismatches  268;  Indels   41;  Gaps   17;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQEG 290
          | || |  |:|||  | |: : |:: :|    |::        ||||  |  |    |  |
Db    137 W-APIPCSMLENSLGPFPLFLQQIQ-SDTAQNYTIYYSIRGPGVDKEPLNLFYVERD-TG 193

Qy    291 DIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQ 350
          ::| |  :|||: :::   | |    |    ||| : :|::| ||| |     :  | ||
Db    194 NLYCTGRVDREQYESFELTAFATTPDGYTPEYPLPLLIKIEDENDNYPIFTQKLYSFTVQ 253

Qy    351 ENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSL 410
          || |:|: :| : | | || :| :: | | |:||:|  ||  || :    |::      |
Db    254 ENSRIGSIVGEVCATDLDEPDTMHTRLRYSILEQSPSPPM--LFTMHPSTGVITTTSAQL 311

Qy    411 KKQDTPQYNLTIEVSDKD-----FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNI 465
           ::   :| | |:| | |       |    | : |:|| :| | :: |  :|  |:  :
Db    312 DRELIDKYQLLIKVQDMDGQYFGLHTTAKCIITIEDVNDNLPTFTRTTY--VTSVEENTV 369

Qy    466 GSTILTIQATDADEPPTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAA 525
             || :   | |   : : :  | |:||:  |   : ||| || | : : ||||:|
Db    370 NVEILRLTVQDKDLVNSPNWRANYTILKGNENGNFKIVTDPKTNEGILCVIKPLDYEERQ 429

Qy    526 VSNIVFKAENPEPLVFGVKYNA-SSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTK 584
             :    | |        : | |  |: ||: :| |:   .:   :: |:| :||:
Db    430 QVTLQIGVVNEAPYTREASSKSPMSTATVTVTVTNQDEGPECIPPMQTVRIQENVPVGTR 489

Qy    585 VGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQV 637
            | |||      | |   | |||: ::  :| |     |||||       | : |
Db    490 NDGYKAYDPETRSSSGIRYRKLSDPRGWVTVNEDSGSITIFRALDREAETVRNGIYNITV 549

Qy    638 VATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHL 697
          :| :  | | :      :|| ||||| | : |    :  |    :   ::    ||
Db    550 LALDADGRSCTGT--LGIILEDVNDNGPFIPKQ--TVVICKATMSSAEIVAVDLDDP--- 602

Qy    698 FRGPHFTFSLGS--GSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLE 755
            || | ||| |    :|  | :::|| | |||| :: |    :| | ||: |  |  |
Db    603 VNGPPFDFSLESSDSEVQRMWRLTRINDTAARLSYQN-DPSFGSYAVPIRVTD--RLGLS 659

Qy    756 GIVSLPVTFCSCV-EGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVF 807
           : :| | | |: |  |   :| :||   | :    :|  || | ::  ::|
Db    660 SVTTLNVLVCDCITESDCTLRSGERTGYADVRLGPWAILAILLGIALLFCILF 712


Search completed: January 15, 2003, 08:15:57
Job time : 23.0457 secs
```

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: January 15, 2003, 08:15:27 ; Search time 56.0365 Seconds
(without alignments)
3059.276 Million cell updates/sec

Title: US-09-079-819-178
Perfect score: 4321
Sequence: 1 MILQAHLHSLCLLMLYLATG.....DKGKDNVESAQASEVKPLRS 832

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2332	54.0	578	11	Q63423	Q63423 rattus norv
2	1427.5	33.0	868	13	Q90X63	Q90x63 brachydanio
3	709.5	16.4	893	13	Q8UVQ7	Q8uvq7 brachydanio
4	699	16.2	922	13	P79883	P79883 xenopus. ma
5	677	15.7	783	13	Q90275	Q90275 brachydanio
6	671.5	15.5	839	4	Q9HAZ9	Q9haz9 homo sapien
7	652	15.1	713	4	Q8TBX3	Q8tbx3 homo sapien
8	652	15.1	714	11	Q8VDK4	Q8vdk4 mus musculu
9	651.5	15.1	814	6	O77704	O77704 canis famil
10	650	15.0	714	11	Q8R490	Q8r490 rattus norv
11	628	14.5	4589	11	Q9WU10	Q9wu10 rattus norv
12	624	14.4	4587	11	Q9QXA3	Q9qxa3 mus musculu
13	611	14.1	4555	11	Q8R508	Q8r508 rattus norv
14	601.5	13.9	3579	5	Q9V5N8	Q9v5n8 drosophila
15	591.5	13.7	4643	5	Q9VW71	Q9vw71 drosophila
16	576	13.3	3014	4	Q9NYQ6	Q9nyq6 homo sapien
17	572.5	13.2	901	4	Q9UII7	Q9uii7 homo sapien
18	571	13.2	3034	11	O35161	O35161 mus musculu
19	570.5	13.2	840	4	Q9HB00	Q9hb00 homo sapien
20	570.5	13.2	894	4	Q9HB01	Q9hb01 homo sapien
21	564	13.1	3301	11	Q91ZI0	Q91zi0 mus musculu
22	561.5	13.0	796	4	Q96CZ9	Q96cz9 homo sapien
23	558	12.9	3313	11	O88278	O88278 rattus norv
24	557.5	12.9	3312	4	Q9NYQ7	Q9nyq7 homo sapien
25	556.5	12.9	788	11	Q8VI68	Q8vi68 mus musculu
26	552	12.8	3503	5	Q24292	Q24292 drosophila
27	545.5	12.6	801	11	Q9Z0M3	Q9z0m3 mus musculu
28	543.5	12.6	693	4	Q15066	Q15066 homo sapien
29	543.5	12.6	792	13	Q9DFS0	Q9dfs0 xenopus lae
30	543.5	12.6	798	13	Q8QGH3	Q8qgh3 gallus gall
31	541.5	12.5	792	13	Q9DFS1	Q9dfs1 xenopus lae
32	538	12.5	864	13	Q90Z37	Q90z37 brachydanio
33	530	12.3	2923	4	Q9HCU4	Q9hcu4 homo sapien
34	528	12.2	2809	5	O61230	O61230 lytechinus
35	521	12.1	4349	4	Q9NYQ8	Q9nyq8 homo sapien
36	518.5	12.0	4351	11	O88277	O88277 rattus norv
37	518	12.0	782	6	Q8WNW5	Q8wnw5 sus scrofa
38	515.5	11.9	910	5	Q95YK9	Q95yk9 ciona savig
39	514.5	11.9	503	4	Q9NTY1	Q9nty1 homo sapien
40	508	11.8	2920	11	Q9R0M0	Q9r0m0 mus musculu
41	496.5	11.5	790	13	Q91838	Q91838 xenopus lae
42	494	11.4	794	13	O93264	O93264 xenopus lae
43	489	11.3	828	4	Q9UJ99	Q9uj99 homo sapien
44	489	11.3	1783	5	Q9VJB6	Q9vjb6 drosophila
45	488.5	11.3	821	4	Q9UII8	Q9uii8 homo sapien

ALIGNMENTS

RESULT 1
Q63423
ID Q63423 PRELIMINARY; PRT; 578 AA.
AC Q63423;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Proton-driven peptide transporter (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=SMALL INTESTINE;
RX MEDLINE=96067558; PubMed=7488096;
RA Erickson R.H., Gum J.R. Jr., Lindstrom M.M., McKean D., Kim Y.S.;
RT "Regional expression and dietary regulation of rat small intestinal
RT peptide and amino acid transporter mRNAs.";
RL Biochem. Biophys. Res. Commun. 216:249-257(1995).
CC -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
DR EMBL; L46874; AAC42077.1; -.
DR HSSP; P15116; 1NCJ.
DR InterPro; IPR002126; Cadherin.
DR Pfam; PF00028; cadherin; 6.
DR SMART; SM00112; CA; 6.
DR PROSITE; PS00232; CADHERIN_1; 1.
DR PROSITE; PS50268; CADHERIN_2; 6.
KW Calcium-binding; Cell adhesion; Glycoprotein.
FT NON_TER 1 1
FT NON_TER 578 578
SQ SEQUENCE 578 AA; 64517 MW; 5644574BC6D14FE0 CRC64;

Query Match 54.0%; Score 2332; DB 11; Length 578;
Best Local Similarity 77.3%; Pred. No. 4.3e-145;
Matches 447; Conservative 55; Mismatches 76; Indels 0; Gaps 0;

Qy 62 LTGETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDIND 121

```
Db      1 LTGETDGILKIEKDGLLYHARVLDRETRAVHHLRLAALDSQGAIVDGPVPIIIEVKDIND 60

Qy    122 NRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQ 181
          |||||||:||||||||||||||||:|||||||||||||||:||||||| ||||||||
Db     61 NRPTFLQTKYEGSVRQNSRPGKPFMYVNATDLDDPATPNGQLFYQIVIQLPKINNVMYFQ 120

Qy    182 INNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKA 241
          |:|||||||| |||| |:| ||| ||||:||||||||:|||||||||| | ||||||
Db    121 IDNKTGAISLTPEGSQVLDPIKNPYYNLVVSVKDMGGQNENSFSDTTSVDITVRENIWKA 180

Qy    242 PKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDRE 301
          |:|||: || ||||||||||:||:| | |||::|||||:|||||||||:|||||||||
Db    181 PEPVEIRENLTDPHPIKITQVQWNEPVAHYSLINKEKLPQFPFSIDQEGNIYVTQPLDRE 240

Qy    302 EKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGT 361
          |||::||:| |||| ||||:||||| ||| |||||||| | ||||||||| ||:|||
Db    241 EKDSHVFFATAKDENGKPLAYPLEIRVKVIDINDNPPTCLSQVTVFEVQENEVLGSSIGI 300

Qy    362 LTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLT 421
            ||| || |  || | ||:|:|||||:| | ||||  | | :|| |:||||||:||||||
Db    301 FAAHDMDEANNINSILMYRLVDQTPKVPSDELFLIDEYGGKVQLGKRSLKKQDSPQYNLT 360

Qy    422 IEVSDKDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPF 481
          :|||| |||||| :|:|||||||||||||:|||||: ||:||| |||||| ||||| ||||
Db    361 VEVSDIDFKTLCSLQVNVIDINDQIPIFERSDYGSKTLSEDTAIGSTILIIQATDDDEPF 420

Qy    482 TGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVF 541
          |||||||| |::||:|||| | ||| |||||| |:|||||||  |::|||||||||||||
Db    421 TGSSKILYKIVQGDTEGRLEVVTDPMTNTGYVKIRKPLDFETEPVTSIVFKAENPEPLVN 480

Qy    542 GVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYS 601
          |::||||||| | | | |||| | | | :||| |||| ||||||| |||:||||| :|||
Db    481 GIEYNASSFASFELTVXDVNEVPVFPQQIFQANVSEDTAIGTKVGTVTARDPEGLTVSYS 540

Qy    602 LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVA 639
          |: : ||||||| |||:||| ||||||  | |||||||
Db    541 LKDNKRGWLKIDSVTGDIFSTAPLDRETESVYRVQVVA 578


RESULT 2
Q90X63
ID   Q90X63      PRELIMINARY;      PRT;   868 AA.
AC   Q90X63;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Cadherin-17.
GN   CDH17.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Horsfield J., Ramachandran A., Reuter K., LaVallie E.,
RA   Collins-Racie L., Crosier K., Crosier P. Sr.;
RT   "Zebrafish cadherin-17 is required for pronephric duct development.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
DR   EMBL; AF428098; AAL29444.1; -.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 6.
DR   PROSITE; PS00232; CADHERIN_1; UNKNOWN_2.
DR   PROSITE; PS50268; CADHERIN_2; 6.
KW   Calcium-binding; Cell adhesion; Glycoprotein.
SQ   SEQUENCE   868 AA;  96677 MW;  40C30844C836E990 CRC64;

  Query Match           33.0%;  Score 1427.5;  DB 13;  Length 868;
  Best Local Similarity   36.5%;  Pred. No. 2.5e-85;
  Matches  320;  Conservative  150;  Mismatches  330;  Indels   77;  Gaps   16;

Qy     13 LMLYLATGYG--QEGKFSGPLKPMTFSIYEGQEPSQIIFQF-KANPPAVTFELTGETDNI 69
          | | :: |:|   | |  |||      : |        |:|  |     :: ::|||::
Db      9 LTLLVSIGHGIDLEDK-KGPLIDTVLDVPEATPVPYAFFKFTSAVEDVSSYRVSGETEDK 67

Qy     70 FVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQS 129
          |  :| ||  : |:      |:| : ||  :|  ::||  : ::| |:|:| | | :|
Db     68 IRISSDGWLYLEQPLEWSPEKKHHLLIEALSEDGKTLDGPAQVVLQVIDVNNNPPVFSES 127

Qy    130 KYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAI 189
          :| ||:|::|  | ||: | |:| ||| | | || : || |:|::    :| ||   | |
Db    128 QYSGSIREHSPAGVPFVQVFASDADDPNTENTQLRFSIVNQIPIVGQTFFFGINPNNGQI 187

Qy    190 SLTREGSQEL---------------------------------------------------- 199
          | ||:: |
Db    188 FTTEEGAEFLKARPSVTYSRGEVRGSPDVLKKKFEDYCIPKNNIALENNPFYKCVERAER 247

Qy    200 ---NPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDPHP 256
             |  ::| | |::  :|:|| : || | || |:| : :|:| :| |: : ||: : :|
Db    248 RTVNVLQDPDYALIVRAEDLGGNAVNSLSSTTRVNIAILQNLWVSPGPITIRENTDEEYP 307

Qy    257 IKITQVRWNDPGAQYSLVDKEKLPRFPFSIDQEGDIYVTQPLDREEKDAYVFYAVAKDEY 316
          : :  :| |:| | | |  || |  |||:|:|:|||||| |||||||: |:   :|:||
Db    308 MYLATMRANNPTALYRLEQKEIL-SFPFTINQDGDIYVTGPLDREEKEMYILVVIAEDEQ 366

Qy    317 GKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSF 376
          |  |  |:|| | |:| |||| |     :|:||| | :|||| | ||| |:| | :|
Db    367 GVELEKPMEIPVLVQDENDNPPMCDE--ALFEVQEKEPVGNSIGHLPAHDNDKEGTLSSA 424

Qy    377 LNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKDFKTLCFVQ 436
          | | :  |||  | | :| |    | :::| |: :::  ||| || ||:|: | | |
Db    425 LTYTLRSQTPTKPSDKMFSIDPNTGEIKVANQNFQRKQVPQYELTFEVTDQVFFTKCKAI 484

Qy    437 INVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDS 496
          | |||||:|||||:|||  :: |   :|:|:| |:|||||:  ||||:: |||  ||
Db    485 IKVIDINDEIPIFEKNDYGTYSVPELAEVGTTLLNIKATDADDTGTGSSRVEYHITAGDP 544

Qy    497 EGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLI 556
          :    :: |  |  | | | :|||:|  :| |:   | |||||: ||:|| ||     :
Db    545 QNLFAIEVDEETGEGRVYIAQPLDYELQSVYNLKIDARNPEPLIAGVEYNDSSTTSVVIE 604

Qy    557 VTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLDISYSLRGDTRGWLKIDHVT 616
          : ||:| |:|        | |:: :|| :    ||||||  | : : ||   ||:::  |
Db    605 LVDVDEPPKFEVEGLNVNVPENITVGTLLMKAEAKDPEGKTIKFKMEGDEHKWLELNVDT 664

Qy    617 GEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFF 676
          ||: : | ||||    : : : | |  || : :  :  : | ||||| |:| |  |  |
Db    665 GELKTKAALDRETVDHFTLTITAYETEGSKMEAEMKVDIHLQDVNDNYPKLQK--TQGFI 722

Qy    677 CHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFE 736
          |     |  |   | | |   : |  |||::   |   ::|:  ::|| |:|  :
Db    723 CLQDMTP--LTLTAMDKDADPY-GEPFTFAISRKS--QNFEIKPVDGTSAKLILKKKPSS 777

Qy    737 ERAYVVLIRINDGGRPPLEGIV-SLPVTFCSCVE-GSCF-RPAGHQTGIPTVGMAVGILL 793
          |:   | | : |       ||     |  |:| : | |:  || |   : ::|  :|||
Db    778 EQNVTVPINVLDNAG---LGITQKFDVRICNCTKLGYCYIEPASHSWKL-SMGSTIGILA 833

Qy    794 TTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKPL 830
            |||: | :   :|||   ||   :    | | :
Db    834 GVFGVIGLFLGICLYQIKK---KDKQRATAEGETKAM 867


RESULT 3
Q8UVQ7
ID   Q8UVQ7      PRELIMINARY;      PRT;   893 AA.
AC   Q8UVQ7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   N-cadherin.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lele Z., Folchert A., Concha M., Rauch G.-J., Geisler R., Rosa F.,
RA   Wilson S.W., Bally-Cuif L., Hammerschmidt M.;
RT   "Zebrafish parachute/n-cadherin is required for neural tube closure
RT   and correct positioning of neurons within the CNS.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF418565; AAL59160.1; -.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   InterPro; IPR000215; Serpin.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; UNKNOWN_3.
DR   PROSITE; PS50268; CADHERIN_2; 3.
DR   PROSITE; PS00284; SERPIN; UNKNOWN_1.
SQ   SEQUENCE   893 AA;  98990 MW;  5E3FF0CDD92B3A6F CRC64;

  Query Match             16.4%;  Score 709.5;  DB 13;  Length 893;
  Best Local Similarity   27.8%;  Pred. No. 4.5e-38;
  Matches  213;  Conservative  115;  Mismatches  335;  Indels  103;  Gaps   21;

Qy    123 RPTFLQSKYEGSVRQNSRPGKPFLYVNATDL----------DDPATPNGQLYYQIVIQLP 172
Db     28 QPGFTENEYNVMTADVITEGQVLLKVDFVDCGRGSGLRFESGDPAD---------------- 73

Qy    173 MINNVMYFQINNKTGAISLTREGSQELNPAKNPSYNLVISVKDMGGQSE----NSFSDTT 228
Db     74 -------FRI-DADGTVMAAR--TLQLTDRKGQS--LEIKAKDENSQEQWMVHINFTQPK 121

Qy    229 SVDIIV----------------TENIWKAPKPVEMVENSTDPHP---IKITQVRWNDPGA 269
Db    122 QVPVILFPRHSVLVKGDDSVNRVKRDWVIP-PVNVLENSRKQFPEELVKIQSDKDKSNTL 180

Qy    270 QYSLV--DKEKLPRFPFSIDQ-EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEI 326
Db    181 RYSVTGPGADQNPTGLFIIDPISGLLSVTKPLDREHIPNFHLRAHAVDINGNQMENPIDI 240

Qy    327 HVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTP 386
Db    241 IINVIDMNDNRPEFTHQIWNGTVDEGAKPGTFVMTVTSQDKDDPNTANGMLRYKILSQTP 300

Qy    387 KLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVI 440
Db    301 ESPSSNMFTINNKTGKIITVAAGLDREKVPQYTLIIQATDMEGNPTYGLSNTATAVIRLL 360

Qy    441 DINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRL 500
Db    361 DVNDNAPEFTRETFHGEVPENRVNVIVTNLTV--TDKDEPGTPAWNAVYRIISGDPTGRF 418

Qy    501 GVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDV 560
Db    419 SIPTDPVTNEGLVTVVKPVDFEMNRSFMLTVVADNEVPLASGIHRTRQSTATVSIRVIDV 478

Qy    561 NEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGL---DISYSLRGDTRGWLKIDHVTG 617
Db    479 NESPNFDPNPKQIKLEEGLPQWSMLTTFTAHDPDRYMQQTISYSKLYDPANWLEIDPNNG 538

Qy    618 EIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTG 673
Db    539 RISTIAVLDRESPYVKNNLYNATFMASDNGVPRASGTGTLQIYLLDINDNAPRV------ 592

Qy    674 LFFCHPLSA-------PGSLIFEATDDDQHLFRGPHFTFSLGS--GSLQNDWEVSKINGT 724
Db    593 ----FPQEAEVCERPEPNAINITAVDGDLNPNAGP-YAFELPNRPSDIRRNWTLTRISGD 647

Qy    725 HARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIP 783
Db    648 HAQLSLKISYLESGIYELPISITDSGNLPMSNTTYLRIKVCQCDHHGDCV----DMERIM 703

Qy    784 TVGMAVGILLTTLLVIGIILAVVFIRIKKDKGKDNVESAQASEVKP 829
```

```
Db    704 AAGLGTGAIIAILICIIILLVLVLMFVMWMKRRDKERQAKQLLIDP 749


RESULT 4
P79883
ID   P79883      PRELIMINARY;      PRT;   922 AA.
AC   P79883;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Maternally expressed neural cadherin XmN-cadherin.
GN   MN-CADHERIN.
OS   Xenopus.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae.
OX   NCBI_TaxID=8353;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TAIL BUD;
RX   MEDLINE=96257964; PubMed=8652409;
RA   Tashiro K., Tooi O., Nakamura H., Koga C., Ito Y., Hikasa H.,
RA   Shiokawa K.;
RT   "Cloning and expression studies of cDNA for a novel Xenopus cadherin
RT   (XmN-cadherin), expressed maternally and later neural-specifically in
RT   embryogenesis.";
RL   Mech. Dev. 54:161-171(1996).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 3 CADHERIN DOMAINS.
DR   EMBL; S82457; AAB37685.2; -.
DR   HSSP; P15116; 1NCI.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 3.
KW   Calcium-binding; Cell adhesion; Glycoprotein.
SQ   SEQUENCE   922 AA;  101173 MW;  FEA7A95CBAF1B640 CRC64;

  Query Match             16.2%;  Score 699;  DB 13;  Length 922;
  Best Local Similarity   26.4%;  Pred. No. 2.3e-37;
  Matches  224;  Conservative  140;  Mismatches  334;  Indels  152;  Gaps   31;

Qy      5 AHLHSLCLLMLYLATGYGQEGKFSGPLKPMTFSIYEGQEPSQIIFQ-FKANPPAVTFELT 63
Db     34 ASVSSAC------TPGFSADG-YTALVSP---NIMEGQKLLKVKFNGCSSGAQGIWYE-T 82

Qy     64 GETDNIFVIEREGLLYYNRALDRETRSTHNLQVAALDANGIIV----EGPVPITIEVKDI 119
Db     83 NNPD--FKVGADGAVY----------TAREVQIPAKQAKFIVAAWDHETP---------- 120

Qy    120 NDNRPTFLQSKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMY 179
Db    121 ---------EKWEAAIQ---------LFVE-------ETPLNQSQYQ------------ 142

Qy    180 FQINNKTGAISLTREGSQE------LNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDII 233
Db    143 ------TG--SQSQESEQEQSQSGTLLPWRQHHKGLRRQKRD----------------- 176

Qy    234 VTENIWKAPKPVEMVENSTDPHPIKITQVRWN---DPGAQYSL--VDKEKLPRFPFSIDQ 288
Db    177 -----WVIP-PVNVPENSRGPFPQQLVLIRSDKDRDDTIRYSITGVGADQPPMAIFNIDP 230

Qy    289 -EGDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVF 347
```

```
Db    231 IFGRMNVTRPLDREERSSYHLRAHAVDINGNKVENPIDLSIYVIDMNDNRPEFSSPIFNG 290

Qy    348 EVQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAK 407
          | |  : |  : |:|||| |: ||:|  : |||::|:|: |   :|:| :  |::
Db    291 SVDEASKPGTYVMTVTAHDADDINTSNGIVMYRIMDQSPQSPSHDMFVIHSKTGVINTVA 350

Qy    408 QSLKKQDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAE 461
            | ::   || : |: :| :                | | |:|| | | :  :  :
Db    351 AGLDREKVQQYTVVIQATDMEGNLNHGLSNTATAIITVADVNDNPPEFTRKMF--IGEVP 408

Qy    462 DTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDF 521
          : ::   :  :   | |:|:| :   :: || || :|   : ||| || | | : ||:|:
Db    409 ENHVDVVVANLTVVDRDQPYTSNWNAVFKIISGDPDGHFTIKTDPVTNEGIVTVSKPVDY 468

Qy    522 ETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAI 581
          | : |  ::    |  ||  |:: :  | |  |: | |||||| |      :  |  :
Db    469 EMSKVFPLIVMVTNQAPLASGIQMSLQSTAAVTVSVNDVNEAPYFPNRNEPIRKLEGESA 528

Qy    582 GTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREAGSP------ 632
          |  :   :| ||:      : ||   |   || |:   |:: : | ||||  ||
Db    529 GRLLITFSAVDPDHSMQQVLRYSKISDPANWLAINTTNGQVSTTAVLDRE--SPFVKDDL 586

Qy    633 YRVQVVATEVGGGSSLSSVSEFHLILMDVNDNPPR-LAKDYTGLFFCHPLSAPGSLIFEAT 691
          |: : :||: |    |      : |:|:||| |  | ||      |   :  | :   |
Db    587 YQAKFLATDNGNPPASGTGTLLIQLIDINDNAPELLPKDAQ---ICERPNGNG-INITAI 642

Qy    692 DDDQHLFRGPHFTFSLGS--GSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDG 749
          | |:     | | | | |   ::: :| : :||  :|||| :    |   | | : : |
Db    643 DVDRKPSADP-FVFELPSVPYTIRRNWTIHRINSIYARLSLQIGYLESGMYDVPVIVTDS 701

Qy    750 GRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVV--F 807
          | |||     :  | |         |    :   |:  | ::: |: | |:|::|  |
Db    702 GNPPLYNTSIIKVKVCPCDNNGDCTTIG---AVAAAGLGTGAIISILICIIILLSMVLLF 758

Qy    808 IRIKKDKGKD 817
          :   | : |:
Db    759 VMWMKRREKE 768


RESULT 5
Q90275
ID   Q90275      PRELIMINARY;      PRT;   783 AA.
AC   Q90275;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Neural-cadherin precursor (N-cadherin).
GN   CDH2 OR CDH2 OR ZNCAD.
OS   Brachydanio rerio (Zebrafish) (Zebra danio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC   TISSUE=EMBRYO;
RX   MEDLINE=95178741; PubMed=7873785;
RA   Bitzur S., Kam Z., Geiger B.;
RT   "Structure and distribution of N-cadherin in developing zebrafish
RT   embryos: morphogenetic effects of ectopic over-expression.";
RL   Dev. Dyn. 201:121-136(1994).
CC   -!- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC       THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC       MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC       SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN
CC       NEURONAL RECOGNITION MECHANISM.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSION IN THE EMBRYO IS INITIALLY FOUND IN
CC       ALL DEEP CELLS BUT LATER BECOMES RESTRICTED TO VARIOUS EPITHELIAL
CC       AND NEURONAL TISSUES. FOUND IN DISCRETE AREAS OF CELL-CELL
CC       ADHESION WHEN EPIBOLY IS NEARLY COMPLETED. EXPRESSED IN ADULT
CC       BRAIN, EYES AND TRUNK.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSION IS FIRST DETECTED IN THE EMBRYO
CC       DURING GASTRULATION.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
DR   EMBL; X67648; CAA47890.1; -.
DR   HSSP; P15116; 1NCI.
DR   ZFIN; ZDB-GENE-990415-171; cdh2.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000233; Cadherin_C_term.
DR   Pfam; PF00028; cadherin; 5.
DR   Pfam; PF01049; Cadherin_C_term; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW   Calcium-binding; Repeat; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?     35       POTENTIAL.
FT   CHAIN        36    783       NEURAL-CADHERIN.
FT   DOMAIN       36    601       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    602    622       POTENTIAL.
FT   DOMAIN      623    783       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       36    143       CADHERIN 1.
FT   REPEAT      144    258       CADHERIN 2.
FT   REPEAT      259    373       CADHERIN 3.
FT   REPEAT      374    480       CADHERIN 4.
FT   REPEAT      481    591       CADHERIN 5.
FT   DOMAIN      740    755       SER-RICH.
FT   CARBOHYD    149    149       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    201    201       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    499    499       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    528    528       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    569    569       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   783 AA;  87120 MW;  B57ECEDF41CCC1C5 CRC64;

  Query Match             15.7%;  Score 677;  DB 13;  Length 783;
  Best Local Similarity   30.1%;  Pred. No. 5e-36;
  Matches  187;  Conservative   98;  Mismatches  288;  Indels   48;  Gaps   15;

Qy    239 WKAPKPVEMVENSTDPHP---IKITQVRWNDPGAQYSLV--DKEKLPRFPFSIDQ-EGDI 292
          |  | || ::|||     |   :||   :       :||:     :: |   | ||   | :
Db     37 WVIP-PVNVLENSRKQFPEELVKIQSDKDKSNTLRYSVTGPGADQNPTGLFIIDPISGLL 95

Qy    293 YVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQEN 352
          ||:|||||     :   | | |  |  :  |::| : | |:||| |     :    | |
Db     96 SVTKPLDREHIPNFHLRAHAVDINGNQMENPIDIIINVIDMNDNRPEFTHQIWNGTVDEG 155

Qy    353 ERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKK 412
          : |  : |:|: |:|: ||||  | |:|: |||: |   :| |    | :      | :
Db    156 AKPGTFVMTVTSQDKDDPNTANGMLRYKILSQTPESPSSNMFTINNKTGKIITVAAGLDR 215

Qy    413 QDTPQYNLTIEVSDKD------FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIG 466
          :  ||| | |: :| :               | ::|:||  | | :  :         |:
Db    216 EKVPQYTLIIQATDMEGNPTYGLSNTATAVIRLLDVNDNAPEFTRETFHGEVPENRVNVI 275

Qy    467 STILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAV 526
          | ||:  || ||| | :   :| || ||  ||  : ||| || | | : ||:|||
Db    276 VTNLTV--TDKDEPGTPAWNAVYRIISGDPTGRFSIPTDPVTNEGLVTVVKPVDFEMNRS 333

Qy    527 SNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVG 586
          :   |:|  ||  |:     | |  :: | ||||:| |  :  | |: | :   : :
Db    334 FMLTVVADNEVPLASGIHRTRQSTATVSIRVIDVNESPNFDPNPKQIKLEEGLPQWSMLT 393

Qy    587 NVTA-KDPEGL---DISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVV 638
          ||  ||:      ||||   |    |:||   | | ::| ||||:      : |    :
Db    394 TFTAHDDPDRYMQQTISYSKLYDPANLLEIDPNNGRISTIAVLDRESPYVKNNLYNATFM 453

Qy    639 ATEVGGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSA-------PGSLIFEAT 691
          |:: |    |      : |:|:||| ||:           |  |       | ::   |
Db    454 ASDNGVPRASGTGTLQIYLLDINDNAPRV----------FPQEAEVCERPEPNAINITAV 503
```

```
Qy    692 DDDQHLFRGPHFTFSLGS--GSLQNDWEVSKINGTHARLSTRHTDFEERAYVVLIRINDG 749
Db    504 DGDLNPNAGP-YAFELPNRPSDIRRNWTLTRISGDHAQLSLKISYLESGIYELPISITDS 562

Qy    750 GRPPLEGIVSLPVTFCSC-VEGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVFI 808
Db    563 GNLPMSNTTYLRIKVCQCDHHGDCV----DMERIMAAGLGTGAIIAILICIIILLVLVLM 618

Qy    809 RIKKDKGKDNVESAQASEVKP 829
Db    619 FVMWMKRRDKERQAKQLLIDP 639


RESULT 6
Q9HAZ9
ID   Q9HAZ9      PRELIMINARY;      PRT;   839 AA.
AC   Q9HAZ9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Desmocollin 3b.
GN   DSC3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20484176; PubMed=11027496;
RA   Whittock N.V., Hunt D.M., Rickman L., Malhi S., Vogazianou A.P.,
RA   Dawson L.F., Eady R.A., Buxton R.S., McGrath J.A.;
RT   "Genomic organization and amplification of the human desmosomal
RT   cadherin genes DSC1 and DSC3, encoding desmocollin types 1 and 3.";
RL   Biochem. Biophys. Res. Commun. 276:454-460(2000).
CC   -!- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
DR   EMBL; AF293359; AAG23427.1; -.
DR   HSSP; P15116; 1NCJ.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 5.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 5.
KW   Calcium-binding; Cell adhesion; Glycoprotein.
SQ   SEQUENCE   839 AA;  93497 MW;  F218E2D8869A6855 CRC64;

  Query Match             15.5%;  Score 671.5;  DB 4;  Length 839;
  Best Local Similarity   30.8%;  Pred. No. 1.3e-35;
  Matches  187;  Conservative  106;  Mismatches  263;  Indels   51;  Gaps   22;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQEG 290
Db    137 W-APIPCSMQENSLGPFPLFLQQVE-SDAAQNYTVFYSISGRGVDKEPLNLFYIERD-TG 193

Qy    291 DIYVTQPLDREEKDAY--VFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFE 348
Db    194 NLFCTRPVDREEYDVFDLIAYASTADGYSADL--PLPLPIRVEDENDNHPVFTEAIYNFE 251

Qy    349 VQENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQ 408
Db    252 VLESSRPGTTVGVVCATDRDEPDTMHTRLKYSILQQTPRSP--GLFSVHPSTGVITTVSH 309

Qy    409 SLKKQDTPQYNLTIEVSDKDFK-------TLCFVQINVIDINDQIPIFEKSDYGNLTLAE 461
Db    310 YLDREVVDKYSLIMKVQDMDGQFFGLIGTSTCI--ITVTDSNDNAPTFRQNAYE--AFVE 365

Qy    462 DTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDF 521
Db    366 ENAFNVEILRIPIEDKDLINTANWRVNFTILKGNENGHFKISTDKETNEGVLSVVKPLNY 425

Qy    522 ETAAVSNIVFKAENPEPLVFGV-KYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVA 580
Db    426 EENRQVNLEIGVNNEAPFARDIPRVTALNRALVTVHVRDLDEGPECTPAAQYVRIKENLA 485

Qy    581 IGTKVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREAGSP----Y 633
Db    486 VGSKINGYKAYDPENRNGNGLRYKKLHDPKGWITIDEISGSIITSKILDREVETPKNELY 545

Qy    634 RVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDD 693
Db    546 NITVLAIDKDDRSCTGTLAVN--IEDVNDNPPEILQEY--VVICKPKMGYTDIL--AVDP 599

Qy    694 DQHLFRGPHFTFSLGSGS--LQNDWEVSKINGTHARLS-TRHTDFEERAYVVLIRINDGG 750
Db    600 DEPVHGAP-FYFSLPNTSPEISRLWSLTKVNDTAARLSYQKNAGFQE--YTIPITVKD-- 654

Qy    751 RPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIPTVGMAV-GILLTTLLVIGIILAVVFIR 809
Db    655 RAGQAATKLLRVNLCECTHPTQCRATSRSTGVILGKWAILAILLGIALLFSVLLTLVCGV 714

Qy    810 IKKDKGK 816
Db    715 FGATKGK 721


RESULT 7
Q8TBX3
ID   Q8TBX3      PRELIMINARY;      PRT;   713 AA.
AC   Q8TBX3;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cadherin 13, H-cadherin (heart).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RA   Strausberg R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC028624; AAH28624.1; -.
SQ   SEQUENCE   713 AA;  78320 MW;  E947F5FA773988FA CRC64;

  Query Match             15.1%;  Score 652;  DB 4;  Length 713;
  Best Local Similarity   28.9%;  Pred. No. 1.9e-34;
  Matches  192;  Conservative  100;  Mismatches  274;  Indels   98;  Gaps   21;

Qy    160 NGQLYYQIVIQLPMINNVMYFQINNKTGAISL---TREGSQELNPAKNPSYN-----LVI 211
Db     61 NDKLRYEV--------SSPYFKVNSDGGLVALRNITAVGKTLFVHARTPHAEDMAELVIV 112

Qy    212 SVKDMGGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVEMVENSTDPHPIKITQVRW 264
Db    113 GGKDIQGSLQDIFKFARTSPVPRQKRSIVVS--------PILIPENQRQPFPRDVGKV-- 162

Qy    265 NDPGAQYSLVDKEKLPRFPF-----SIDQE----------GDIYVTQPLDREEKDAYVF 308
Db    163 ---------VDSDRPERSKFRLTGKGVDQEPKGIFRINENTGSVSMTRTLDREVIAVYQL 213

Qy    309 YAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRD 368
Db    214 FVETTDVNGKTLEGPVPLEVIVIDQNDNRPIFREGPYIGHVMEGSPTGTTVMRMTAFDAD 273

Qy    369 EENTANSFLNYRIVEQTPKLPMDGLFLIQTYAG-MLQLAKQSLKKQDT---PQYNLTIEV 424
Db    274 DPATDNALLRYNIRRQTPDKPSPNMFYIDPEKGDIVTVVSPALLDRETLENPKYELIIEA 333

Qy    425 SDK-----DFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADE 479
Db    334 QDMAGLDVGLTGTATATIMIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDD 390

Qy    480 PFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPL 539
Db    391 PATGAWRAAYTIINGNPGQSFEIHTNPQTNEGMLSVVKPLDYEISAFHTLLIKVENEDPL 450
```

```
Qy    540 VFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLD-- 597
Db    451 VPDVSYGPSSTATVHITVLDVNEGPVFYPDPMMVTRQEDLSVGSVLLTVNATDPDSLQHQ 510

Qy    598 -ISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSE 652
Db    511 TIRYSVYKDPAGWLNINPINGTVDTTAVLDRESTFVDNSVYTALFLAIDSGNPPATGTGT 570

Qy    653 FHLILMDVNDNPPRLAKDYTGLFFCHPLSAP--------GSLIFEATDDDQHLFRGPHFT 704
Db    571 LLITLEDVNDNAP----------FIYPTVAEVCDDAKNLSVVILGASDKDLHPNTDP-FK 619

Qy    705 FSLGSGSLQND-WEVSKINGTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVT 763
Db    620 FEIHKQAVPDKVWKISKINNTHALVSLLQ-NLNKANYNLPIMVTDSGKPPMTNITDLRVQ 678

Qy    764 FCSC 767
Db    679 VCSC 682


RESULT 8
Q8VDK4
ID   Q8VDK4      PRELIMINARY;      PRT;   714 AA.
AC   Q8VDK4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cadherin 13.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC021628; AAH21628.1; -.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 5.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 5.
DR   PROSITE; PS00232; CADHERIN_1; UNKNOWN_3.
DR   PROSITE; PS50268; CADHERIN_2; 3.
SQ   SEQUENCE   714 AA;  78116 MW;  CA5CA791C46A2FFB CRC64;

  Query Match           15.1%;  Score 652;  DB 11;  Length 714;
  Best Local Similarity   29.3%;  Pred. No. 1.9e-34;
  Matches  189;  Conservative  109;  Mismatches  287;  Indels   60;  Gaps   20;

Qy    160 NGQLYYQIVIQLPMINNVMYFQINNKTGAISL---TREGSQELNPAKNPSYN-----LVI 211
Db     61 NEKLHYEV--------SSPHFKVNSDGTLVALRNITAVGRTLFVHARTPHAEDMAELVIV 112

Qy    212 SVKDMGGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVEMVENSTDPHPIKITQVRW 264
Db    113 GGKDIQGSLQDIFKFARTSPVPRQKRSIVVS--------PILIPENQRQPFPRDVGKVVD 164

Qy    265 ND--PGAQYSLVDK--EKLPRFPFSIDQE-GDIYVTQPLDREEKDAYVFYAVAKDEYGKP 319
Db    165 SDRPEGSKFRLTGKGVDQDPKGTFRINENTGSVSVTRTLDRETIATYQLYVETTDASGKT 224

Qy    320 LSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNY 379
Db    225 LEGPVPLEVIVIDQNDNRPIFREGPYIGHVMEGSPTGTTVMRMTAFDADDPATDNALLRY 284

Qy    380 RIVEQTPKLPMDGLFLIQTYAG-MLQLAKQSLKKQDT---PQYNLTIEVSDK-----DFK 430
Db    285 NIRQQTPDKPSPNMFYIDPEKGDIVTVVSPALLDRETLENPKYELIIEAQDMAGLDVGLT 344

Qy    431 TLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYH 490
Db    345 GTATATIVIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDDPTTGAWRAAYT 401

Qy    491 IIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSF 550
Db    402 IINGNPGQSFEIHTNPQTNEGMLSVVKPLDYEISAFHTLLIKVENEDPLVPDVSYGPSST 461

Qy    551 AKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLD---ISYSLRGDTR 607
Db    462 ATVHITVLDVNEGPVFYPDPMMVTKQENISVGSVLLTVNATDPDSLQHQTIRYSVYKDPA 521

Qy    608 GWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDN 663
Db    522 GWLSINPINGTVDTTAVLDRESPFVHNSVYTALFLAIDSGNPPATGTGTLLITLEDINDN 581

Qy    664 PPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQND-WEVSKIN 722
Db    582 APVIYP--TVAEVCDDARNLSVVILGASDKDLHPNTDP-FKFEIHKQTVPDKVWKISKIN 638

Qy    723 GTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC 767
Db    639 NTHALVSLLQ-NLNKANYNLPIMVTDSGKPPMTNITDLRVQVCSC 682


RESULT 9
O77704
ID   O77704      PRELIMINARY;      PRT;   814 AA.
AC   O77704;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Desmocollin type 2 (Fragment).
GN   DSC2.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98380123; PubMed=9716266;
RA   Roberts G.A., Burdett I.D., Pidsley S.C., King I.A., Magee A.I.,
RA   Buxton R.S.;
RT   "Antisense expression of a desmocollin gene in MDCK cells alters
RT   desmosome plaque assembly but does not affect desmoglein expression.";
RL   Eur. J. Cell Biol. 76:192-203(1998).
CC   -!- SIMILARITY: CONTAINS 4 CADHERIN DOMAINS.
DR   EMBL; AJ002299; CAA05309.1; -.
DR   HSSP; P09803; 1SUH.
DR   InterPro; IPR002126; Cadherin.
DR   Pfam; PF00028; cadherin; 5.
DR   SMART; SM00112; CA; 4.
DR   PROSITE; PS00232; CADHERIN_1; 2.
DR   PROSITE; PS50268; CADHERIN_2; 4.
KW   Calcium-binding; Cell adhesion; Glycoprotein.
FT   NON_TER       1      1
FT   CHAIN       101    814       DESMOCOLLIN TYPE 2.
SQ   SEQUENCE   814 AA;  91081 MW;  59EF086C69BD3882 CRC64;

  Query Match           15.1%;  Score 651.5;  DB 6;  Length 814;
  Best Local Similarity   30.2%;  Pred. No. 2.5e-34;
  Matches  181;  Conservative  110;  Mismatches  256;  Indels   53;  Gaps   20;

Qy    239 WKAPKPVEMVENSTDPHPIKITQVRWNDPGAQYSL--------VDKEKLPRFPFSIDQE- 289
Db    102 W-APIPCSMQENSLGPFPLFLQQIQ-SDTAQNYTIFYSIRGPGVDRE--PKNLFYVERDT 157

Qy    290 GDIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEV 349
Db    158 GNLFCTRPVDREEYESFELIAFATTPDGYTPELPLPLVIRIEDENDNYPIFTEKTYVFTI 217

Qy    350 QENERLGNSIGTLTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQS 409
Db    218 SENCRVGSTVGQVCATDKDEPDTMHTRLKYSIIEQLPAYPT--LFSMHPATGVITTSSSQ 275
```

```
Qy    410 LKKQDTPQYNLTIEVSDKD-----FKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTN 464
Db    276 LDRELIDKYQLKIKVQDMDGQYFGLQTTSICIINIDDVNDNLPTFTRTSY--VTSVEENT 333

Qy    465 IGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETA 524
Db    334 VDVEILRVTVEDKDLMNTANWRANYTILKGNENENFKIVTDPKTNEGILCVVKPLNYEER 393

Qy    525 AVSNIVFKAENPEPL-VFGVKYNASSFAKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGT 583
Db    394 KQVDLQIVVVNEAPYSKEGSLRSTMSTATVTVNVKNQDEGPECSPRVQTIQTKENVPVGT 453

Qy    584 KVGNVTAKDPE---GLDISYSLRGDTRGWLKIDHVTGEIFSVAPLDREA----GSPYRVQ 636
Db    454 EIQGYKAYDPETRSSSGIRYKKLSDPKEWVRINENTGSITIFRNLDREAEMIRGGIYNIT 513

Qy    637 VVATEVGGSSLSSVSEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQH 696
Db    514 ILASDKDGRSCTGT--LGIILQDVNDNGPSIPKH--TVVICKTVLSPSADIV-AIDPDEP 568

Qy    697 LFRGPHFTFSLGS---GSLQNDWEVSKINGTHARLSTRHTDFEER-----AYVVLIRIND 748
Db    569 I-NGPPFDFRLENIPDSDIERTWRLTKINDTAARL------FYEKDLPFGTYRISVRVAD 621

Qy    749 GGRPPLEGIVSLPVTFCSCV-EGSCFRPAGHQTGIPTVGMAVGILLTTLLVIGIILAVVF 807
Db    622 --RHGRSLITPLKVLLCDCVTENDCTLRTDARTGGGDVRLGKWAILAILLGIALLFCILF 679


RESULT 10
Q8R490
ID   Q8R490      PRELIMINARY;      PRT;   714 AA.
AC   Q8R490;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   T-cadherin.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HEART;
RA   Niermann T., Schmutz S.;
RT   "Cloning of rat T-cadherin.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF494095; AAM14607.1; -.
SQ   SEQUENCE   714 AA;  78085 MW;  7E56142B8162F864 CRC64;

  Query Match             15.0%;  Score 650;  DB 11;  Length 714;
  Best Local Similarity   29.1%;  Pred. No. 2.6e-34;
  Matches  188;  Conservative  110;  Mismatches  287;  Indels   60;  Gaps   20;

Qy    160 NGQLYYQIVIQLPMINNVMYFQINNKTGAI---SLTREGSQELNPAKNPSYN-----LVI 211
Db     61 NKKLHYEV--------SSPYFKVNSDGTLVAVRNITAVGRTLFVHARTPHAEDMAELVIV 112

Qy    212 SVKDMGGQSEN--SFSDTTSV-----DIIVTENIWKAPKPVEMVENSTDPHPIKITQVRW 264
Db    113 GGKDIQGSLQDIFKFARTSPVPRQKRSIVVS--------PILIPENQRQPFPRDVGKVVD 164

Qy    265 ND--PGAQYSLVDK--EKLPRFPFSIDQE-GDIYVTQPLDREEKDAYVFYAVAKDEYGKP 319
Db    165 SDRPEGSKFRLTGKGVDQDPKGTFRINENTGSVSVTRTLDRETIATYQLFVETTDASGRT 224

Qy    320 LSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTLTAHDRDEENTANSFLNY 379
Db    225 LEGPVPLEVIVIDQNDNRPIFREGPYIGHVMEGSPTGTTVMRMTAFDADDPATDNALLRY 284

Qy    380 RIVEQTPKLPMDGLFLIQTYAG-MLQLAKQSLKKQDT---PQYNLTIEVSDK-----DFK 430
Db    285 NIRQQTPDKPSPNMFYIDPEKGDIVTVVSPALLDRETLENPKYELIIEAQDMAGLDVGLT 344

Qy    431 TLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYH 490
Db    345 GTATATIVIDDKNDHSPKFTKKEF--QATVEEGAVG-VIVNLTVEDKDDPATGAWRAAYT 401

Qy    491 IIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSF 550
Db    402 IINGNPGQSFEIHTNPQTNEGMLSVVKPLDYEISAFHTLLIKVENEDPLVPDVSYGPSST 461

Qy    551 AKFTLIVTDVNEAPQFSQHVFQAKVSEDVAIGTKVGNVTAKDPEGLD---ISYSLRGDTR 607
Db    462 ATVHITVLDVNEGPVFYPDPMMVTKQENISVGSVLLTVNATDPDSLQHQTIRYSVYKDPA 521

Qy    608 GWLKIDHVTGEIFSVAPLDREA----GSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDN 663
Db    522 GWLSINPINGTVDTTAVLDRESPFVHNSVYTALFLAIDSGNPPATGTGTLMITLEDVNDN 581

Qy    664 PPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQND-WEVSKIN 722
Db    582 APVIYP--TVAEVCDDARNLSVVILGASDKDLHPNTDP-FKFEIHKQTVPDKVWKISKIN 638

Qy    723 GTHARLSTRHTDFEERAYVVLIRINDGGRPPLEGIVSLPVTFCSC 767
Db    639 NTHALVSLLQ-NLNKANYNLPIMVTDSGKPPMTNITDLRVQVCSC 682


RESULT 11
Q9WU10
ID   Q9WU10      PRELIMINARY;      PRT;  4589 AA.
AC   Q9WU10;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Protocadherin.
GN   FAT.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY;
RX   MEDLINE=99173791; PubMed=10072790;
RA   Ponassi M., Jacques T.S., Ciani L., ffrench Constant C.;
RT   "Expression of the rat homologue of the Drosophila fat tumour
RT   suppressor gene.";
RL   Mech. Dev. 80:207-212(1999).
CC   -!- SIMILARITY: CONTAINS 33 CADHERIN DOMAINS.
DR   EMBL; AF100960; AAD20459.1; -.
DR   HSSP; P01132; 1EGF.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00028; cadherin; 33.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 31.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 2.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00232; CADHERIN_1; 16.
DR   PROSITE; PS50268; CADHERIN_2; 33.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_4.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Calcium-binding; Cell adhesion; EGF-like domain; Glycoprotein;
KW   Hydroxylation; Repeat.
SQ   SEQUENCE   4589 AA;  505987 MW;  8CD0561B1DC58677 CRC64;
```

```
  Query Match             14.5%;  Score 628;  DB 11;  Length 4589;
  Best Local Similarity   25.3%;  Pred. No. 1.3e-31;
  Matches  228;  Conservative  135;  Mismatches  329;  Indels  210;  Gaps   32;

Qy      26 KFSGPLKPMTFSIYE----GQEPSQIIFQ---------FKANPPAVTFELTGETDNIFVI 72
           :|| |    |:|| |    | |   |  :          | | |           |  |||
Db    2703 RFSEPF--YTYSISEDMPIGTEIDLIRVEHSGTVLYTLVKGNTP------ESNRDEFFVI 2754

Qy      73 ERE-GLLYYNRALDRETRSTHNLQVAA---LDANGIIVEGPVPITIEVKDINDNRPTFLQ 128
           :|: | |   ::|| ||   :   : |   ||   ::    : ::|:||| ||| |
Db    2755 DRQSGRLKLEKSLDHETTKWYQFSILARCTLDDYEVVAS--IDVSIQVKDANDNSPVLES 2812

Qy     129 SKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGA 188
           : ||  : :|   |   : | |:|||     |||: | :  |    : :  | || :||
Db    2813 NPYEAFIVENLPAGSRVIQVRASDLDSGV--NGQVMYSL-DQSQDADIIESFAINMETGW 2869

Qy     189 ISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE----------NI 238
           |:  :    ||:  :  || : :   | | : :   | |  ||: ||:           |
Db    2870 ITTLK----ELDHEERASYQIKVVASDHGEKVQ--LSSTAIVDVTVTDVNDSPPRFTAEI 2923

Qy     239 WKA-------PKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPF-SIDQEG 290
           :|        |  |  : ::||    :| :         | :   : | :|   ::  :
Db    2924 YKGTVSEDDPPGGVIAILSTTDADSEEINR------QVSYFITGGDALGQFAVENMQNDW 2977

Qy     291 DIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTC---------- 340
            :|| :|||||:||:|:    | |      |    : ||| | ||| | |
Db    2978 RVYVKKPLDREQKDSYLLTVTATD---GTFSSKARVEVKVLDANDNSPVCERTAYSDAIP 3034

Qy     341 ------------------------------------------------------------ 340

Db    3035 EDALPGKLVMQVSATDADIRSNAEITYTLFGSGAEKFKLNPDTGELRTLALLDREEQAVY 3094

Qy     341 ----------------------------------PSPVTVFEVQENERLGNSIGTLTAH 365
                                              |  :|||   ||   |  :  : |
Db    3095 HLLVKATDGGGRSCQATIVLTLEDVNDNTPEFTADPYAITVF---ENTEPGTPLTRVQAT 3151

Qy     366 DRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVS 425
           | |     |  ::| :::       || | |   :|:||| |  | ::    | ||:: :
Db    3152 DADA--GLNRKISYSLLDSA-----DGQFSINEQSGILQLEKH-LDRELQAVYTLTLKAA 3203

Qy     426 D----KDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPF 481
           |    :       | ::|:|||  |:||  :||  |::||  ||: :| : |   |
Db    3204 DQGLPRKLTATGTVVVSVLDINDNPPVFEYREYG-ATVSEDIVIGTEVLQVYAASRD--I 3260

Qy     482 TGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVF 541
             :::| | || |:  |:  :|    : || : | : ||:|::    :  :| :
Db    3261 EANAEITYAIISGNEHGKFSID----SKTGAIFIIENLDYESSHGYYLTVEATD------ 3310

Qy     542 GVKYNASSFAKFTLIVTDVNE-APQFSQHVFQAKVSEDVAIGTKVGNVTAKD---PEGLD 597
           |   : |  |   : :||:|: :| |||  :   ||||  |:   |  : | |   |
Db    3311 GGTPSLSDVATVNINITDINDNSPVFSQDSYTTVVSEDAALEQPVITIMADDADGPSNSH 3370

Qy     598 ISYS-LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLI 656
           | || : |:      || | |||    |||||  | | : | | : |     :  ::
Db    3371 ILYSIIEGNQGSPFTIDPVRGEIKVTKPLDRETISGYTLTVQAADNGNPPRVNTTTVNID 3430

Qy     657 LMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDW 716
           : ||||| |  ::|   :          |::     |      || |:|:: ||:  | :
Db    3431 VSDVNDNAPLFSRDNYSVIIQENKPVGFSVLKLVVTDKDSSHNGPPFSFAIVSGNDDNMF 3490

Qy     717 EVSKINGTHARLSTRHTDFE--ERAYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFR 774
           ||::    |  | |  |     :  |:: ::: | |:| |  :  : :     :| |
Db    3491 EVNQ----HGVLLTAATVKRKVKDHYLLHVKVADNGKPQLSSLTHIDI---RVIEESIHP 3543

Qy     775 PA 776
           ||
Db    3544 PA 3545


RESULT 12
Q9QXA3
ID   Q9QXA3      PRELIMINARY;      PRT;  4587 AA.
AC   Q9QXA3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Mouse fat 1 cadherin (Fragment).
GN   MFAT1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EMBRYO;
RX   MEDLINE=20203461; PubMed=10741417;
RA   Cox B.T.M., Hadjantonakis A.K., Collins J., Magee A.I.;
RT   "Cloning and expression throughout mouse development of mfat1 a
RT   homologue of the Drosophila tumour suppressor gene fat.";
RL   Dev. Dyn. 217:233-240(2000).
CC   -!- SIMILARITY: CONTAINS 33 CADHERIN DOMAINS.
DR   EMBL; AJ250768; CAB65271.1; -.
DR   HSSP; P08709; 1BF9.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00028; cadherin; 33.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 31.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00232; CADHERIN_1; 16.
DR   PROSITE; PS50268; CADHERIN_2; 33.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_2.
KW   Calcium-binding; Cell adhesion; Glycoprotein.
FT   NON_TER    4587   4587
SQ   SEQUENCE   4587 AA;  506036 MW;  4D3F23BB05127CB4 CRC64;

  Query Match             14.4%;  Score 624;  DB 11;  Length 4587;
  Best Local Similarity   25.1%;  Pred. No. 2.5e-31;
  Matches  227;  Conservative  135;  Mismatches  329;  Indels  212;  Gaps   33;

Qy      26 KFSGPLKPMTFSIYE----GQEPSQIIFQ---------FKANPPAVTFELTGETDNIFVI 72
           :|| |    |::| |    | |   |  :          | | |           |  |||
Db    2704 RFSEPF--YTYTISEDTPIGTEIDLIRVEHGGAVLYILVKGNTP------ESNRDEFFVI 2755

Qy      73 ERE-GLLYYNRALDRETRSTHNLQVAA---LDANGIIVEGPVPITIEVKDINDNRPTFLQ 128
           :|: | |   ::|| ||   :   : |   ||   ::    : ::|:||| ||| |
Db    2756 DRQNGRLKLEKSLDHETTKWYQFSILARCTLDDYEVVAS--IDVSIQVKDANDNSPVLES 2813

Qy     129 SKYEGSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGA 188
           | ||  : :|   |   : : |:|||  |  |||: | :  |    : :  | || :||
Db    2814 SPYEAFIVENLPGGSRVIQIRASDLDSGA--NGQVMYSL-DQSQDADIIESFAINMETGW 2870

Qy     189 ISLTREGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTE----------NI 238
           |:  :    ||:  :  || : :   | | : :   | |  | : ||:           |
Db    2871 ITTLK----ELDHEERASYQIKVVASDHGEKVQ--LSSTAIVGVTVTDVNDSPPRFTAEI 2924

Qy     239 WKA-------PKPVEMVENSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRFPF-SIDQEG 290
           :|        |  |  : ::||    :| :         | :   : | :|   ::  :
Db    2925 YKGTVSEDDPPGGVIAILSTTDADTEEINR------QVSYFITGGDALGQFAVENVQSDW 2978

Qy     291 DIYVTQPLDREEKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTC---------- 340
            :|| :|||||:||:|:    | |      |    : ||| | ||| | |
Db    2979 RVYVKKPLDREQKDSYLLTVTATD---GTFSSKARVEVKVLDANDNSPVCEKTSYSDTIP 3035

Qy     341 ------------------------------------------------------------ 340

Db    3036 EDALPGKLVMQVSATDADIRSNAEITYTLFGSGAEKFKLNPDTGELRTLALLDREEQAVY 3095

Qy     341 ----------------------------------PSPVTVFEVQENERLGNSIGTLTAH 365
```

```
Db   3096 NLLVKATDGGGRSCQAAIVLTLEDVNDNAPEFTAEPYTITVF---ENTEPGTPLTRVQAT 3152

Qy    366 DRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVS 425

Db   3153 DAD--TGLNRKISYSLVESA-----DGQFSINERSGIIQLEKH-LDRELQAVYTLTLKAV 3204

Qy    426 D----KDFKTLCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPF 481

Db   3205 DQGLPRRLTATGTVVVSVLDINDNPPVFEYREYG-ASVSEDIVIGTEVLQVYAASRD--I 3261

Qy    482 TGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVF 541

Db   3262 EANAEITYAIISGNEHGKFSID----SKTGAIFIIESLDYESSHEYYLTVEATD------ 3311

Qy    542 GVKYNASSFAKFTLIVTDVNE-APQFSQHVFQAKVSEDVAIGTKVGNVTAKD---PEGLD 597

Db   3312 GGTPSLSDVATVNINVTDINDNSPVFSQDTYTTVVSEDAALEQPVITIMADDADGPSNSH 3371

Qy    598 ISYS-LRGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLI 656

Db   3372 IHYSIIEGNQGSPFTIDPVRGEVKVTKPLDRETISGYTLTVQAADNGNPPRVNTTTVNID 3431

Qy    657 LMDVNDNPPRLAKDYTGLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDW 716

Db   3432 VSDVNDNAPLFSRDNYSVIIQENKPVGFSVLKLVVTDKDSSHNGPPFFFTIVSGNDENAF 3491

Qy    717 EVSKINGTHARLSTRHTDFEERA---YVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCF 773

Db   3492 EVNQ----HGVLLTAAT-IKRKVKDHYFLHVKVADSGKPQLSSMTHIDI---RVIEESIH 3543

Qy    774 RPA 776

Db   3544 PPA 3546


RESULT 13
Q8R508
ID   Q8R508      PRELIMINARY;      PRT;  4555 AA.
AC   Q8R508;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Fat3.
GN   FAT3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX   MEDLINE=21670969; PubMed=11811999;
RA   Mitsui K., Nakajima D., Ohara O., Nakayama M.;
RT   "Mammalian fat3: a large protein that contains multiple cadherin and
RT   EGF-like motifs.";
RL   Biochem. Biophys. Res. Commun. 290:1260-1266(2002).
DR   EMBL; AB076401; BAB86869.1; -.
SQ   SEQUENCE   4555 AA;  502077 MW;  B47CBC10638EA73B CRC64;

  Query Match           14.1%;  Score 611;  DB 11;  Length 4555;
  Best Local Similarity   25.4%;  Pred. No. 1.7e-30;
  Matches  225;  Conservative  144;  Mismatches  320;  Indels  196;  Gaps   33;

Qy     30 PLKPMTFSIYEGQEPSQIIFQFKANPPAVTFELTGETDNIFVIERE-GLLYYNRALDRET 88

Db   2733 PNQSVRFSTVNGERPE-----------------NNKENVFIIEQETGAIKLDKRLDHEV 2774

Qy     89 RSTHNLQVAA---LDANGIIVEGPVPITIEVKDINDNRPTFLQSKYEGSVRQNSRPGKPF 145

Db   2775 SPAFHFKVAATIPLDKVDIVF--TVDVDVKVLDLNDNKPVFETSSYETIIMEGMPVGTKL 2832

Qy    146 LYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTREGSQELNPAKNP 205

Db   2833 AQVRAIDTDWGA--NGQVTYSLHSDSHLEKVMEAFNIDSNTGWISTLK----DLDHETDP 2886

Qy    206 SYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPKPVEMVENSTDPHPIKITQVRWN 265

Db   2887 TFSFFVVASDLG--EAFSLSSMALVSVKVTDINDNAP---------VFAHEVYRGNVKES 2935

Qy    266 DPG----AQYSLVDKEKL--------------PRFPFS---IDQEGDIYVTQPLDREEKD 304

Db   2936 DPPGEVVAVLSTLDKDTSNINRQVSYHITGGNPRGRFALGMVQSEWKVYVKRPLDREEQD 2995

Qy    305 AYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTC------------PSPVTVFEVQE 351

Db   2996 IYFLNITASD--GLFVTQAM-VEVTVSDVNDNSPVCDQVAYSASLPEDIPSNKIILKVSA 3052

Qy    352 NER--------------LGNS----------IGTLTAHDRDE----------------- 369

Db   3053 KDADIGSNGDIRYSLYGSGNSDFFLDPESGELKTLALLDRERVPVYNLIARATDGGGRFC 3112

Qy    370 ----------------------------ENTANSFLNYRIVEQTPKLPMD--------- 391

Db   3113 SSTVLLLLEDVNDNPPVFSSNHYTACVYENTATKALLTRVQAVDPDVGINRKVVYSLEDS 3172

Qy    392 --GLFLIQTYAGMLQLAKQSLKKQDTPQYNLTIEVSD----KDFKTLCFVQINVIDINDQ 445

Db   3173 ASGVFSIDSSSGVIVL-EQPLDREQQSSYNISVRATDQSPGQSLSSLTSVTITVLDINDN 3231

Qy    446 IPIFEKSDYGNLTLAEDTNIGSTILTIQATDADEPFTGSSKILYHIIKGDSEGRLGVDTD 505

Db   3232 PPVFERRDY-LVTVPEDTSLGTQVLSVFATSKD--IGTNAEITYLIRSGNEQGKFRI--- 3285

Qy    506 PHTNTGYVIIKKPLDFETAAVSNIVFKAENPEPLVFGVKYNASSFAKFTLIVTDVNE-AP 564

Db   3286 -NPKTGGISVLEALDYEMCKRFYLVVEAKD------GGTPALSTAATVSIDLTDVNDNPP 3338

Qy    565 QFSQHVFQAKVSEDVAIGTKVGNVTAKD----PEGLDISYSL-RGDTRGWLKIDHVTGEI 619

Db   3339 RFSQDVYSAVISEDALEGDSVILLIAEDVDSKPNG-QIRFSIVGGDRDNEFAVDPILGLV 3397

Qy    620 FSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRLA-KDYTGLFFCH 678

Db   3398 KVKKKLDRERVSGYSLLIQAVDSGIPAMSSTTTVNIDISDVNDNSPVFTPANYTAVIQEN 3457

Qy    679 PLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGSLQNDWEVSKINGTHARLSTRHTDFEER 738

Db   3458 KPVGTSILQLVVTDRDS-FHNGPPFSFSILSGNEDEEFMLDSHGILRSAVVFRHMESPE- 3515

Qy    739 AYVVLIRINDGGRPPLEGIVSLPVTFCSCVEGSCFRPAGHQTGIP 783

Db   3516 -YLLCIQAKDSGKPQQ---VSHTYIRVRVIEESTHKP----TAIP 3552


RESULT 14
Q9V5N8
ID   Q9V5N8      PRELIMINARY;      PRT;  3579 AA.
AC   Q9V5N8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PROTOCADHERIN-like wing polarity protein STAN precursor (STARRY NIGHT
DE   protein) (FLAMINGO protein).
GN   STAN OR CG11895 OR FMI.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=EMBRYO;
RX   MEDLINE=20025940; PubMed=10556066;
RA   Chae J.W., Kim M.-J., Goo J.H., Collier S., Gubb D., Charlton J.,
RA   Adler P.N., Park W.J.;
```

```
RT   "The Drosophila tissue polarity gene starry night encodes a member of
RT   the protocadherin family.";
RL   Development 126:5421-5429(1999).
RN   [2]
RP   SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX   MEDLINE=99418630; PubMed=10490098;
RA   Usui T., Shima Y., Shimada Y., Hirano S., Burgess R.W., Schwarz T.L.,
RA   Takeichi M., Uemura T.;
RT   "Flamingo, a seven-pass transmembrane cadherin, regulates planar cell
RT   polarity under the control of frizzled.";
RL   Cell 98:585-595(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: INVOLVED IN THE FZ SIGNALING PATHWAY THAT CONTROLS WING
CC       TISSUE POLARITY. ALSO MEDIATES HOMOPHILIC CELL ADHESION. MAY PLAY
CC       A ROLE IN INITIATING PREHAIR MORPHOGENESIS.
CC   -!- SUBCELLULAR LOCATION: AT 12 HOURS AFTER PUPARIUM FORMATION (APF),
CC       EXPRESSED EVENLY AT CELL BOUNDARIES. BY 30 HOURS APF, EXPRESSION
CC       IS CONCENTRATED AT PROXIMAL AND DISTAL CELL BOUNDARIES WITH LITTLE
CC       OR NO EXPRESSION AT ANTERIOR AND POSTERIOR BOUNDARIES. WHEN
CC       PREHAIRS EMERGE AT 30-36 HOURS APF, EXPRESSION BECOMES EVENLY
CC       DISTRIBUTED AGAIN ALONG THE WHOLE CELL BOUNDARY.
CC   -!- TISSUE SPECIFICITY: IN THE PUPAL WING, EXPRESSED AT RELATIVELY
CC       EVEN LEVELS IN ALL REGIONS.
CC   -!- DEVELOPMENTAL STAGE: ABUNDANT IN 6-9 HOUR EMBRYOS. EXPRESSED AT
CC       HIGHER LEVELS IN PUPAE THAN LARVAE.
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 LAMININ EGF-LIKE DOMAIN.
DR   EMBL; AF172329; AAF02618.1; -.
DR   EMBL; AB028498; BAA84069.1; -.
DR   EMBL; AE003828; AAF58763.2; ALT_SEQ.
DR   HSSP; P08709; 1BF9.
DR   FlyBase; FBgn0024836; stan.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR000832; GPCR_secretin.
DR   InterPro; IPR001879; hormn_receptor.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00002; 7tm_2; 1.
DR   Pfam; PF00028; cadherin; 8.
DR   Pfam; PF00008; EGF; 3.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF02793; HRM; 1.
DR   Pfam; PF00053; laminin_EGF; 1.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00249; GPCRSECRETIN.
DR   SMART; SM00112; CA; 8.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 4.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00008; HormR; 1.
DR   SMART; SM00282; LamG; 2.
DR   PROSITE; PS00232; CADHERIN_1; 5.
DR   PROSITE; PS50268; CADHERIN_2; 8.
DR   PROSITE; PS00022; EGF_1; 4.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS50227; G_PROTEIN_RECEP_F2_3; 1.
DR   PROSITE; PS50261; G_PROTEIN_RECEP_F2_4; 1.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 1.
KW   Cell adhesion; Repeat; Laminin EGF-like domain; EGF-like domain;
KW   Signal; Glycoprotein; Calcium-binding.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   3579       PROTOCADHERIN-LIKE WING POLARITY PROTEIN
FT                                STAN.
FT   REPEAT      368    464       CADHERIN 1.
FT   REPEAT      473    581       CADHERIN 2.
FT   REPEAT      590    689       CADHERIN 3.
FT   REPEAT      698    794       CADHERIN 4.
FT   REPEAT      803    897       CADHERIN 5.
FT   REPEAT      906   1007       CADHERIN 6.
FT   REPEAT     1016   1113       CADHERIN 7.
FT   REPEAT     1122   1220       CADHERIN 8.
FT   DOMAIN     1482   1518       EGF-LIKE 1.
FT   DOMAIN     1756   1792       EGF-LIKE 2.
FT   DOMAIN     1965   2000       EGF-LIKE 3.
FT   DOMAIN     2091   2126       EGF-LIKE 4.
FT   DOMAIN     1584   1753       LAMININ G-LIKE 1.
FT   DOMAIN     1826   1963       LAMININ G-LIKE 2.
FT   DOMAIN     2092   2140       LAMININ EGF-LIKE.
FT   DISULFID   1486   1497       POTENTIAL.
FT   DISULFID   1491   1506       POTENTIAL.
FT   DISULFID   1508   1517       POTENTIAL.
FT   DISULFID   1760   1771       POTENTIAL.
FT   DISULFID   1765   1780       POTENTIAL.
FT   DISULFID   1782   1791       POTENTIAL.
FT   DISULFID   1969   1979       POTENTIAL.
FT   DISULFID   1973   1988       POTENTIAL.
FT   DISULFID   1990   1999       POTENTIAL.
FT   DISULFID   2092   2095       POTENTIAL.
FT   DISULFID   2097   2114       POTENTIAL.
FT   DISULFID   2116   2125       POTENTIAL.
FT   CARBOHYD     46     46       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    179    179       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    340    340       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    671    671       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    886    886       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1269   1269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1374   1374       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD   1441   1441       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1650   1650       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1678   1678       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1747   1747       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1843   1843       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1975   1975       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2016   2016       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2028   2028       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2071   2071       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2088   2088       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2196   2196       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2320   2320       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2784   2784       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3083   3083       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3228   3228       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    181    181       T -> S (IN REF. 1).
FT   CONFLICT    361    361       Q -> L (IN REF. 1).
FT   CONFLICT    395    401       MVSLLDS -> NGLTVGLP (IN REF. 2).
FT   CONFLICT   1968   1968       Q -> H (IN REF. 1).
FT   CONFLICT   2271   2271       G -> E (IN REF. 1).
FT   CONFLICT   2502   2502       R -> C (IN REF. 2).
FT   CONFLICT   2627   2627       D -> G (IN REF. 2).
FT   CONFLICT   2709   2709       T -> S (IN REF. 2).
FT   CONFLICT   2756   2756       Q -> R (IN REF. 2).
FT   CONFLICT   2901   2901       C -> Y (IN REF. 1).
FT   CONFLICT   3098   3098       L -> P (IN REF. 1).
FT   CONFLICT   3569   3579       ERNIDDDETTV -> DSEAEY (IN REF. 2).
SQ   SEQUENCE   3579 AA;  397139 MW;  4E801C493031FB19 CRC64;

  Query Match           13.9%;  Score 601.5;  DB 5;  Length 3579;
  Best Local Similarity  29.1%;  Pred. No. 5e-30;
  Matches  224;  Conservative  131;  Mismatches  321;  Indels   95;  Gaps   35;

Qy     37 SIYEGQEPSQIIFQFKANP------PAVTFELTGETDN----------IFVIE-REGLLY 79
          || ||      :   :|          : : :   ||           || |: | |::
Db    472 SIREGATVGSTVITLRATDQDIGKNAEIEYGIEAVTDGAGLAQDQEMPIFRIDSRSGVIS 531

Qy     80 YNRALDRETRSTHNLQVAALDANGIIVE---GPVPITIEVKDINDNRPTFLQSKYEGSVR 136
          :|||||  :::| | | |      |       : ::| | ||| | | :  |    |
Db    532 TRSSLDRETSDSYHLLVTAADLASAQSERRTATASVQVKVLDDNDNYPQFSERTYTVQVP 591

Qy    137 QNSRPG---KPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLTR 193
          ::   |       :: ||| |     |  : | |:      |    | |:: :| :|| :
Db    592 EDQWGGTEDNTVAHIRATDADQ--GNNAAIRYAIIGG----NTQSQFSIDSMSGDVSLVK 645

Qy    194 EGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTENIWKAPK------PVEM 247
              |:     || |||  :| |  |    |:|| : : | :    ||:          :
Db    646 ----PLDYESVRSYRLVIRAQDGGSPSR---SNTTQLLVNVIDANDNAPRFYTSQFQESV 698

Qy    248 VENSTDPHPIKITQVRWNDPGAQ----YSLVDKEKLPRFPFSID-QEGDIYVTQPLDREE 302
          :||    : |   |   :| ||     ||: :::    || ::| : | :   :||||||
Db    699 LENVPVGYNIIRVQAYDSDEGANAEITYSISERD--DNFPLAVDPRTGWVQTIKPLDREE 756

Qy    303 KDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGTL 362
          :  : |  ||||    | |    : : |:|:||| |          | |::  |  : |:
Db    757 QGRFAFQVVAKDGGVPPKSASSSVVITVQDVNDNDPAFNPKYYEANVGEDQPPGTPVTTV 816

Qy    363 TAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLI--QTYAGMLQLAKQSLKKQDTPQYNL 420
          || | ||:    | |:| |      :    | | |  |   |:: :| |||  :   :: |
Db    817 TATDPDED----SRLHYEITTGNTR----GRFAITSQNGRGLITIA-QSLDYKQEKRFLL 867

Qy    421 TIEVSDKDFKT-LCFVQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDADE 479
          |:  :|   ::    | ||: | |:  |||| : | : :: ||  :|:|:|: : |||:|
Db    868 TVAATDSGGRSDTATVHINITDANNFAPIFENAPY-SASVFEDAPVGTTVLVVSATDSD- 925

Qy    480 PFTGSSKILYHIIKGDSEGRLGVDTDP---HTNTGYVIIKKPLDFETAAVSNIVFKAENP 536
             | :  : : :  :|   ||   ||   :  || ::   ||| || :   :   |::
Db    926 --VGVNAQITYSLNEESINGLG-SPDPFSINPQTGAIVTNAPLDRETTSGYLLTVTAKD- 981

Qy    537 EPLVFGVKYNASSFAKFTLIVTDVNE-APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPE- 594
             |   : |      : |||||: || |   ::|| : ||  :|| |  | | ||:
Db    982 -----GGNPSLSDTTDVEIGVTDVNDNAPAFKSPLYQASILEDALVGTSVIQVAASDPDV 1036

Qy    595 GLD--ISYSL--RGDTRGWLKIDHVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSV 650
          ||:  | | |  |    |   ||  :| | :   ||||: : : :  :| : |   |||
Db   1037 GLNGRIKYLLSDRDIEDGSFVIDPTSGTIRTNKGLDRESVAVFHLTAIAVDKGSPPLSST 1096

Qy    651 SEFHLILMDVNDNPPRLAKDYTGLFFCHPLSAP-GSLIFE--ATDDDQHLFRGPHFTFSL 707
             |  : | ||||:||  | |   |:   | ::| ||:: |  | | |: :    |  :|:
Db   1097 VEVQIRLEDVNDSPPTFASDKITLYV--PENSPVGSVVGEIHAHDPDEGVNAVVH--YSI 1152

Qy    708 GSGSLQNDWE-VSKINGTHARLST-RHTDFE--ERAYVVLIRINDGGRPPL 754
             |   | :  |::     |:| |    |:|   : : :::|      |||
Db   1153 IGGDDSNAFSLVTRPGSERAQLLTMTELDYESTRKRFELVVR---AASPPL 1200


RESULT 15
Q9VW71
ID   Q9VW71      PRELIMINARY;      PRT;  4643 AA.
AC   Q9VW71;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative fat-like cadherin precursor (CG7749 protein).
GN   CG7749.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- FUNCTION: MAY BE INVOLVED IN CELL ADHESION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC   -!- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
```

```
CC   -!- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAIN.
DR   EMBL; AE003515; AAF49078.1; -.
DR   HSSP; P15116; 1NCI.
DR   FlyBase; FBgn0036930; fat2.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR002126; Cadherin.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   Pfam; PF00028; cadherin; 31.
DR   Pfam; PF00008; EGF; 5.
DR   Pfam; PF00054; laminin_G; 1.
DR   PRINTS; PR00205; CADHERIN.
DR   SMART; SM00112; CA; 33.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 5.
DR   SMART; SM00282; LamG; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00232; CADHERIN_1; 17.
DR   PROSITE; PS50268; CADHERIN_2; 41.
DR   PROSITE; PS00022; EGF_1; 5.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Hypothetical protein; Cell adhesion; Signal; Transmembrane; Repeat;
KW   EGF-like domain; Glycoprotein; Calcium-binding.
FT   SIGNAL        1     35       POTENTIAL.
FT   CHAIN        36   4643       PUTATIVE FAT-LIKE CADHERIN.
FT   DOMAIN       36   1647       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1648   1668       POTENTIAL.
FT   DOMAIN     1669   4643       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       68    180       CADHERIN 1.
FT   DOMAIN      189    288       CADHERIN 2.
FT   DOMAIN      293    397       CADHERIN 3.
FT   DOMAIN      406    504       CADHERIN 4.
FT   DOMAIN      513    610       CADHERIN 5.
FT   DOMAIN      619    713       CADHERIN 6.
FT   DOMAIN      778    874       CADHERIN 7.
FT   DOMAIN      883    977       CADHERIN 8.
FT   DOMAIN      986   1085       CADHERIN 9.
FT   DOMAIN     1094   1190       CADHERIN 10.
FT   DOMAIN     1199   1296       CADHERIN 11.
FT   DOMAIN     1305   1402       CADHERIN 12.
FT   DOMAIN     1413   1503       CADHERIN 13.
FT   DOMAIN     1512   1609       CADHERIN 14.
FT   DOMAIN     1618   1714       CADHERIN 15.
FT   DOMAIN     1723   1812       CADHERIN 16.
FT   DOMAIN     1821   1929       CADHERIN 17.
FT   DOMAIN     1948   2030       CADHERIN 18.
FT   DOMAIN     2024   2137       CADHERIN 19.
FT   DOMAIN     2146   2232       CADHERIN 20.
FT   DOMAIN     2247   2338       CADHERIN 21.
FT   DOMAIN     2347   2465       CADHERIN 22.
FT   DOMAIN     2474   2567       CADHERIN 23.
FT   DOMAIN     2576   2670       CADHERIN 24.
FT   DOMAIN     2679   2779       CADHERIN 25.
FT   DOMAIN     2788   2876       CADHERIN 26.
FT   DOMAIN     2885   2983       CADHERIN 27.
FT   DOMAIN     2992   3083       CADHERIN 28.
FT   DOMAIN     3092   3185       CADHERIN 29.
FT   DOMAIN     3194   3289       CADHERIN 30.
FT   DOMAIN     3298   3394       CADHERIN 31.
FT   DOMAIN     3403   3499       CADHERIN 32.
FT   DOMAIN     3508   3604       CADHERIN 33.
FT   DOMAIN     3618   3671       CADHERIN 34.
FT   DOMAIN     3819   3879       EGF-LIKE 1.
FT   DOMAIN     3881   3919       EGF-LIKE 2.
FT   DOMAIN     4129   4166       EGF-LIKE 3.
FT   DOMAIN     4168   4205       EGF-LIKE 4.
FT   DOMAIN     4243   4279       EGF-LIKE 5.
FT   DISULFID   3823   3835       POTENTIAL.
FT   DISULFID   3830   3867       POTENTIAL.
FT   DISULFID   3869   3878       POTENTIAL.
FT   DISULFID   3885   3896       POTENTIAL.
FT   DISULFID   3890   3907       POTENTIAL.
FT   DISULFID   3909   3918       POTENTIAL.
FT   DISULFID   4133   4144       POTENTIAL.
FT   DISULFID   4138   4154       POTENTIAL.
FT   DISULFID   4156   4165       POTENTIAL.
FT   DISULFID   4172   4183       POTENTIAL.
FT   DISULFID   4177   4193       POTENTIAL.
FT   DISULFID   4195   4204       POTENTIAL.
FT   DISULFID   4247   4258       POTENTIAL.
FT   DISULFID   4252   4267       POTENTIAL.
FT   DISULFID   4269   4278       POTENTIAL.
FT   CARBOHYD     65     65       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    156    156       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    364    364       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    779    779       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    843    843       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    923    923       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1106   1106       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1198   1198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1312   1312       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1439   1439       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1473   1473       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1511   1511       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   4643 AA;  517614 MW;  7EC52BD36ADD4E89 CRC64;

  Query Match           13.7%;  Score 591.5;  DB 5;  Length 4643;
  Best Local Similarity  24.2%;  Pred. No. 3.4e-29;
  Matches  210;  Conservative  153;  Mismatches  312;  Indels  193;  Gaps   33;

Qy      33 PMTFSIYEGQ------EPSQIIFQ-FKANPPAVT------------FELTGETDNIFVI 72
           |::  | |          | | :: :  :: ||                | :  : |: |:|
Db    2764 PVSIEIIETDANIPTFEKSSVLLKIIESTPPGTVLTKLHMIGNYTFKFSIAADQDH-FMI 2822

Qy      73 EREGLLYYNRALDRETRSTHNLQVAALDANGIIVEGPVPITIEVKDINDNRPTFLQSKYE 132
             | |   : |||| : :||| | |  :   :      : |:|:| ||| | |  : |
Db    2823 SDSGELILQQTLDREQQESHNLIVVAETSTVPVFFAYADVLIDVRDENDNYPKFDNTFYS 2882

Qy     133 GSVRQNSRPGKPFLYVNATDLDDPATPNGQLYYQIVIQLPMINNVMYFQINNKTGAISLT 192
            || :||      : |:||| |    ||| : | :      | |:  | |:  :| |:|
Db    2883 ASVAENSEKVISLVKVSATDAD--TGPNGDIRYYLESDTENIQNI--FDIDIYSGWITLL 2938

Qy     193 REGSQELNPAKNPSYNLVISVKDMGGQSENSFSDTTSVDIIVTEN--IWKAP-KPVEMVE 249
             :|:       ||  :   | |    ::    |   :   :|  ::| | : : : |
Db    2939 TSLDREV----QSEYNFKVIAADNGHPKHDAKVPVTIKIVDYNDNAPVFKLPIEGLSVFE 2994

Qy     250 NSTDPHPIKITQVRWNDPGAQYSLVDKEKLPRF--------PFSIDQEGDIYVTQPLDRE 301
           |:  |  : |  :   ||      ::|::  |          | | : |:::: :|||||
Db    2995 NAL-PGTVLI-NLLLIDPD-----IEKQEMDFFIVSGDKQAQFQIGKSGELFIAKPLDRE 3047

Qy     302 EKDAYVFYAVAKDEYGKPLSYPLEIHVKVKDINDNPPTCPSPVTVFEVQENERLGNSIGT 361
           :   |    :| |  ||  :    : : |||||| | |  |       |:  :| ::
Db    3048 QLMFYNLSIIATD--GK-FTAKANVEIDVKDINDNTPYCLKPRYHISTNESISIGTTLVE 3104

Qy     362 LTAHDRDEENTANSFLNYRIVEQTPKLPMDGLFLIQTYAGMLQLAKQSLKKQDTPQYNLT 421
           : | | | ::    :|: :  :          | |   :|:|::|  :| :: ||:| |
Db    3105 VKAIDFDFQSKLRFYLSGKGADD---------FSIGKESGILKVA-SALDRETTPKYKLV 3154

Qy     422 IEVSD-KDFKTLCF--VQINVIDINDQIPIFEKSDYGNLTLAEDTNIGSTILTIQATDAD 478
             | |||   ||  : | | |||| :|||  : |  ::: ||  : : |  : | | |
Db    3155 AHVQDGKDFTQECFSEIIITVNDINDNMPIFSMAQY-RVSVPEDAQLNTLITKVHAMDKD 3213

Qy     479 EPFTGSSKILYHIIKGDSEGRLGVDTDPHTNTGYVIIKKPLDFETAAVSNIVFKAE---- 534
              | :: : : : |::     :      :|| : : | || || :: |:  |||
Db    3214 ---FGVNRQIKYSLMGENHDYFKIS----KSTGIIRLHKSLDRETISLFNLTVKAEDCGV 3266

Qy     535 ----------------NPEPLVFGVK---------------------------------- 544
                           | | | ::
Db    3267 PKLHSIATVAVNILDINDNPPEFSMRQYSCKILENATHGTEVCKVYATSIDIGVNADIHY 3326

Qy     545 --------------------------YNASSF-----------------AKFTLIVTD 559
                                      | | |                 |   : : |
Db    3327 FIMSGNEQGKFKMDSTTGDLVLNATLDYEMSKFYFLTIQAIDGGTPPLSNNAYVNISILD 3386
```

```
Qy     560 VNE-APQFSQHVFQAKVSEDVAIGTKVGNVTAKDPE----GLDISYSL-RGDTRGWLKID 613
           :|: :| | |:::: |:||: :|:|: :| | | :    || ::|:: ||| |   ||
Db    3387 INDNSPTFLQNLYRINVNEDIFVGSKILDVKATDEDSDVNGL-VTYNIERGDNIGQFSID 3445

Qy     614 HVTGEIFSVAPLDREAGSPYRVQVVATEVGGSSLSSVSEFHLILMDVNDNPPRL-AKDYT 672
             | |    |||||  | | ::: | : |     :    :: ::| ||| |   : :|:
Db    3446 PKNGTISVSRPLDRETISHYTLEIQACDQGDPQRCNSVPININILDTNDNAPIFSSSNYS 3505

Qy     673 GLFFCHPLSAPGSLIFEATDDDQHLFRGPHFTFSLGSGS------LQNDWEVSKINGTHA 726
            :   : |     | |: :| |:     | :|| : ||:      |: |  :
Db    3506 VVLQENRLLGYVFLTFKISDADETPNTTP-YTFDIRSGNEGGLFRLEQDGSL-------- 3556

Qy     727 RLSTRHTDFEERAYVVLIRINDGGRPPL 754
           | ::|     :  :|: :|: | | |||
Db    3557 RTASRFNHNLQDEFVIQVRVFDNGTPPL 3584
```

Search completed: January 15, 2003, 08:19:45
Job time : 76.0365 secs